(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
15 November 2001 (15.11.2001)

PCT

(10) International Publication Number
WO 01/85962 A1

(51) International Patent Classification[7]: C12N 15/62, 15/70, C07K 1/30

(21) International Application Number: PCT/GB01/02043

(22) International Filing Date: 4 May 2001 (04.05.2001)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:

0001666-7	5 May 2000 (05.05.2000)	SE
60/209,331	2 June 2000 (02.06.2000)	US
0002432-3	28 June 2000 (28.06.2000)	SE

(71) Applicant *(for all designated States except US)*: **IMPERIAL COLLEGE INNOVATIONS LIMITED** [GB/GB]; Sherfield Building, Imperial College, London SW7 2AZ (GB).

(71) Applicants and
(72) Inventors: **IWATA, So** [JP/GB]; Flat 5, 72 Pelham Road, London SW19 1NY (GB). **BYRNE, Bernadette** [GB/GB]; Flat 6, 7 Embankment Gardens, London SW3 4LG (GB). **JORMAKKA, Mika** [SE/GB]; Flat 2, 87 Brondesbury Villas, London NW3 6AG (GB). **ABRAMSON, Jeff** [US/SE]; Studentvagen 17:22, S-752 34 Uppsala (SE). **SEJLITZ, Torsten** [SE/SE]; St Eriksgatan 59, S-112 34 Stockholm (SE).

(74) Agent: **MILES, John, S.**; Eric Potter Clarkson, Park View House, 58 The Ropewalk, Nottingham NG1 5DD (GB).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published:
— *with international search report*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

WO 01/85962 A1

(54) Title: CRYSTALLOGRAPHY METHODS

(57) Abstract: The present invention provides a recombinant vector comprising, (i) a promoter sequence and (ii) a nucleotide sequence encoding a first protein which, when crystallized with a second protein, is capable of accommodating the second protein in the crystal lattice; said recombinant vector further allowing for the insertion of a further nucleotide sequence encoding a second protein to be located, when crystallized, in the crystal lattice of the first protein. The invention further provides a recombinant vector comprising (i) a promoter sequence and (ii) a nucleotide sequence encoding a first protein which upon crystallization yields crystals having available space in the lattice, so as to allow for the ordered packing of a second protein into the said available space, said recombinant vector further allowing, for the insertion of a further nucleotide sequence encoding a second protein to be accommodated, upon its crystallization, in the said available space in the lattice of the first protein.

B2

CRYSTALLOGRAPHY METHODS

TECHNICAL FIELD

The present invention relates to protein crystallography methods and constructs useful therein, in particular to fusion proteins comprising a first protein and a second protein, whereby the first protein upon crystallization yields crystals having available space in the lattice, so as to allow for the ordered packing of the second protein into the said available space. The invention also relates to methods of crystallization of such a second protein, which e.g. can be a protein, such as a membrane protein, which is otherwise difficult to crystallize. The invention further relates to i.a. recombinant vectors adapted for the expression of a fusion protein as described above.

BACKGROUND ART

Membrane proteins are involved in a multitude of biological processes; the respiratory chain, photosynthesis, transport of solute molecules and ions and regulating cellular responses to a wide range of biological molecules such as hormones, neurotransmitters and drugs. High resolution structural data has allowed useful insight into the function of a number of integral membrane proteins including the G-protein coupled receptor (GPCR), rhodopsin (Palczewski *et al*, 2000), the bc_1 complex (Iwata *et al* 1998), bacteriorhodopsin (Pebay-Peyroula *et al*, 1997) and cytochrome c oxidase (Iwata *et al*, 1995, Tsukihara *et al*, 1996). Despite these successes the number of resolved membrane protein structures remains extremely small compared with soluble proteins.

Crystallization is necessary to obtain the three-dimensional structure of proteins; it often represents the bottleneck in structure determination. In particular, crystallization of membrane proteins has been difficult (For

reviews, see e.g. Durbin & Feher (1996) Annual Review of Physical Chemistry 47, 171-204; and Garavito *et al.* (1996) Journal of Bioenergetics & Biomembranes 28, 13-27). To date, of approximately 12,000 protein structures deposited in the Brookhaven Protein Database, only 20 are membrane proteins. Furthermore, of these 20, only two are eukaryotic in origin (Iwata *et al.* (1998) *Science* 281, 64-71; and Tsukihara *et al.* (1996) *Science* 272, 1136-1144). These are tiny numbers in light of the fact that it is estimated that 35-40% of the genes within the human genome code for integral membrane or membrane associated proteins. Low levels of endogenous expression, high hydrophobicity and low stability in solution all combine to frustrate the membrane protein crystallographer. In many cases, even if enough pure protein can be obtained, it is impossible to grow crystals suitable for structural analysis if indeed they grow at all. In addition, many proteins fail to produce in *E. coli*-based expression systems, which have a number of advantages over other expression systems, including low cost and rapid production of high cell densities.

Progress in solving some of these problems has been made using fusion protein technology. Fusion of soluble proteins, such as maltose binding peptide, to the hydrophobic neurotensin receptor, a GPCR, has been shown to increase expression and facilitate purification (Tucker & Grisshammer, 1996). In addition, the crystallisation of cytochrome c oxidase was facilitated by the addition of a monoclonal F_v fragment (Ostermeier *et al*, 1995). Alternatively, crystallisation in lipidic cubic phase has yielded a high-resolution structure of bacteriorhodopsin (Landau & Rosenbusch, 1996; Pebay-Peyroula *et al*, 1997). However, widely applicable methods to facilitate the crystallisation of membrane proteins have yet to be described.

Membrane protein crystals often contain a very high solvent content (65-80%; Abramson and Iwata (1999), in *Protein Crystallisation*, Ed. Terese

Bergfors, International University Line pp199-210). This solvent space is filled mainly with detergent micelles and can form very large gaps within the crystal lattice structure, gaps which we have found are large enough to accommodate other proteins.

Carter *et al* (1994), in *Protein and Peptide Lett.* 1:175 used a fusion between a six amino acid fragment of a HIV polypeptide and glutathione S-transferase (GST) to allow crystallisation and x-ray analysis of the HIV fragment. In this case, the HIV fragment formed an extension to the GST, forming an integral part of the GST structure.

In Privé (1994) *Acta Cryst.* D50:375-379 and Privé and Kaback (1996) *J. Bioenergetics Biomembranes* 28:29-34, the cytochrome b_{562} was cloned within one of the extracellular loops of the LacP protein. This only acted to extend the hydrophilic domains of the highly hydrophobic LacP protein, and the fusion was not suitable for x-ray analysis since it was incapable of producing suitable crystals.

A similar approach was used by Iwata *et al* (1995) *Nature* 376:660 where an antibody fragment was used to allow crystallisation of cytochrome c oxidase.

However, the GST, cytochrome b_{562} "carrier" molecules and antibody fragments used previously only provide an extra soluble domain, and none are suitable, for example, by being relatively large enough, to accommodate a second protein or protein fragment within their crystal lattice structure or the available space of the crystal.

The cytochrome *bo*3 ubiquinol oxidase from *E. coli* is a member of the heme-copper superfamily of proton-pumping respiratory oxidases (for a

review of the heme-copper respiratory oxidases, see García-Horsman *et al.* (1994) *J. Bacteriol.* 176, 5587-5600). Cytochrome *bo*3 *is* a four-subunit respiratory enzyme (Fig. 1) that catalyses the four-electron reduction of 0_2 to water and functions as a proton pump (Puustinen *et al.* (1991) *Biochemistry* 30, 3936-).

The genes for the cytochrome *bo*3 subunits are organized within a single operon called the *cyo* operon (Chepuri *et al.* (1990) *J. Biol. Chem.* 265, 11185-11192), which is under control of a constitutive, multicistronic promoter (Minacgawa *et al.* (1990) *J. Biol. Chem.* 265, 11198-11203). The sequence of the *cyo* operon and the amino acid sequences of the subunits are disclosed in Chepuri et al. *(supra)* and in GenBank with accession number J05492 (SEQ ID NO: 13). The *cyo* operon has been shown to encode five open reading, frames, cyoABCDE (cf. SEQ ID NOS: 1 and 13). The gene products of *cyoa, cyob, cvoc* and *cyod* correspond to the cytochrome bo3 subunits II, I, III and IV, respectively. The *cyoE* gene encodes a protoheme IX farnesyltransferase (Saiki et al. (1993) *Biochem. Biophys. Res. Comm.* 189, 1491-1497).

Purification of histidine-tagged cytochrome bo3 from *E. coli,* using Ni^{2+} afinity chromatography, is disclosed by Rumbley *et al.* (1997) Biochimica et Biophysica Acta 1340, 131-142).

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1

Structure of cytochrome *bo*3. The upper drawing, shows a periplasmic view of the protein, showing the position of subunit IV relative to the other transmembrane spanning helices. The heme group is shown in the upper right hand corner. The lower drawing shows the same protein looking through the membrane.

Figure 2

Two-dimensional model of subunit IV of cytochrome *bo*3, showing the three transmembrane spanning domains, intracellular N-terminus and extracellular C-terminus

Figure 3

Wire model of cytochrome *bo*3. Subunit IV is shown in the lower left hand corner with the three transmembrane spanning helices labeled. The space available within the crystal lattice (>100 kDa) is shown adjacent.

Figure 4

The cytochrome *bo*3 fusion vectors. Expression of the separate subunits is under control of a single multi-cistronic promoter. A multiple cloning site was introduced at the 3'end of subunit IV and this was used to clone in linkers, to act as bridge sequences, and the proteins of interest.

Figure 5

Western blot analysis of membranes prepared from GO105 cells expressing pMB908 (UBO only), pMB930, pMB946 and pMB947. Panel A shows a blot probed with anti-his antibody directed against the His9 tag at the C-terminal end of subunit II. This blot clearly shows the expression of all of

the modified UBO vector constructs (subunit II is 33 kDa) and that this expression can be localized to the membrane. Panel B shows a blot probed with anti-HA antibody directed against the HA tag in the linker in pMB947. A clear band is seen corresponding to the 17 kDa subunit which is not present for the untagged control.

Figure 6

Crystals of native cytochrome *bo*3, cytochrome *bo*3 + protein Z and cytochrome bo3 + apo A-I. The crystals have different forms, with the native cytochrome *bo*3 forming rod like crystals (diffract to 3.5 Å) while the cytochrome *bo*3 + protein Z crystals (diffract to 6 Å) form as square-plates. Crystals of cytochrome *bo*3 + apo A-I (diffract to 5 Å) form elongated hexagonal plates.

Figure 7

Expression of cytochrome *bo*3 + GPCR fusion proteins. This blot was probed with anti-HA antibody directed against the HA tag introduced at the C-terminal end of subunit IV of cytochrome *bo*3 as a linker sequence. Lane 1 shows undetectable signal for cells only. Lane 2 shows the positive control of *bo*3 + linker only, which gives a band of 14 kDa corresponding to subunit IV only. Lanes 3 and 4 show the expression of subunit IV + either M1 or CB2 receptor. Faint signals can be seen corresponding to the full-length fusion in both cases.

Figure 8

Expression of cytochrome bo3 + leader peptide/ProW constructs. The blot has been probed with an anti-His antibody and shows the specific (33 kDa) bands corresponding to the His-tagged subunit II of cytochrome *bo*3 for the

control, native *bo*3, and the two fusion proteins. No band is present for cells only.

Figure 9

Expression of cytochrome *bo*3 + Apo AI constructs. Each sample was grown to an OD600 of 1.0 prior to harvest. Blot A was probed with Apo AI-specific antibody and shows both whole cells and membranes for the cytochrome *bo*3 + Apo AI constructs. Lane 1 shows the + ve control of a fragment of pure human Apo AI and lane 2 the -ve control of the native cytochrome *bo*3 construct (pMB908) with no detectable signal. Cytochrome *bo*3 + Apo AI (pMB1241) yields two specific bands, a full-length product (30 kDa) and a breakdown product (18 kDa). For cytochrome *bo*3 + *Strep*-tag + Apo AI (pMB1242) only one full-length band is detected, while no specific bands are detected for cytochrome *bo*3 + *Strep-HA*-tag, + Apo AI (pMB1243). Cytochrome *bo*3 + protein Z + Apo AI (pMB1244) exhibits the highest level of expression although this fusion protein also undergoes a certain amount of breakdown, yielding a full-length product (50 kDa) and a smaller fragment (14 kDa). The same pattern of expression is seen for both whole cells and membranes demonstrating the localization of the fusion protein to the membrane. Blot B (lanes 1-8) was probed with anti-His antibody specific for the His9 tag at the C-terminal end of subunit II. No signal is detected for the -ve control of cells only. The +ve control of cytochrome *bo*3 shows a distinct band corresponding to subunit II (33 kDa). All the other constructs yield similar bands except pMB1243 (not shown on this blot) and interestingly all seem to exhibit higher expression levels than the control sample. Lane 9 of blot B shows pMB1244 probed with peroxidase anti-peroxidase specific to protein Z. Interestingly, it is only possible to detect the breakdown product using this conjugate, no full length protein can be observed.

Figure 10

Effects of arabinose concentration on the expression of cytochrome *bo*3 + pBAD. The blot has been probed with anti-His antibody specific for the His-tag, at the C-terminal end of subunit II. The cells containing the cytochrome *bo*3 + pBAD were grown to the mid-log phase and then induced with increasing concentrations of arabinose as detailed on the gel. Cells were harvested 4 h post-induction. No signals are detected for either GL101 cells only or for the uninduced control while a strong band is detected for the +ve control, cytochrome *bo*3 under control of the native constitutive promoter. The production of cytochrome *bo*3 under control of the inducible promoter showed a clear dose response relationship with increasing concentrations of arabinose yielding increasing levels of detectable protein up to 0.2% arabinose. The amount of protein expressed at 0.2% arabinose was much higher than that under control of the native constitutive promoter. However, arabinose concentrations higher than 0.2% had no further increase on the expression level (results not shown). The results are the same for whole cells and membranes deinonstrating the localization of the protein to the membrane.

Figure 11

Effects of induction time on the expression of cytochrome *bo*3 + pBAD promoter. The blot is probed with anti-His antibody specific to the His-tag at the C-terminal end of subunit II. A distinct band is seen for the control of native *bo*3 while no signal is detected for the uninduced cytochrome *bo*3 + pBAD. After the cells had reached the mid-log phase protein production was induced with 0.2% arabinose and the cells incubated at +37°C for the times indicated on the gel prior to sampling. A strong signal is detected one hour post-induction with a slight increase thereafter to 3 h after induction. After this time there is a steady drop in the detectable expression of the tagged protein, to about a 50% reduction after the cells had been incubated

overnight (18 h). It is unclear whether this loss in detectable protein is a result of degradation of the whole cytochrome bo3 molecule or specific proteolysis of the His-tag.

Figure 12

Crystal packing for cytochrome *bo*3 + apo A-I. Cytochrome *bo*3 is shown in white/pale grey and apo A-I in grey.

DISCLOSURE OF THE INVENTION

In a first aspect, this invention provides a recombinant vector comprising (i) a promoter sequence and (ii) a nucleotide sequence encoding a first protein which is a membrane protein or multisubunit protein and which, when crystallized with a second protein, is capable of accommodating the second protein in the crystal lattice; said recombinant vector further allowing for the insertion of a further nucleotide sequence encoding a second protein to be located, when crystallized, in the crystal lattice of the first protein wherein the resulting crystal lattice is capable of diffracting x-rays.

By "crystal lattice" we mean the crystal lattice produced by the first protein.

It will be appreciated however, that the crystal lattice is formed by the protein which makes and maintains most of the crystal contacts within the lattice, and that the crystal lattice itself may be altered by the presence of a second protein. Such an altered crystal lattice is included in our definition of "crystal lattice".

A second aspect of the invention provides a recombinant vector comprising (i) a promoter sequence and (ii) a nucleotide sequence encoding a first protein which is a membrane protein or multisubunit protein and which upon crystallization yields crystals having available space in the lattice, so as to allow for the ordered packing of a second protein into the said available space; said recombinant vector further allowing for the insertion of a further nucleotide sequence encoding a second protein to be accommodated, upon its crystallization, in the said available space in the lattice of the first protein wherein the resulting crystal lattice is capable of diffracting x-rays.

The said "space" may be utilized to force a second "target" protein to pack in an ordered manner into the crystal lattice of the first protein, which is used as a "scaffold" molecule. In addition, fusing a second protein to a first protein may facilitate the expression, folding stability in *E. coli* of the said second protein. The recombinant vectors according to the invention thus provide a template for facilitated/improved expression, purification, crystallization and subsequent structure determination of proteins.

In a preferred embodiment of the first and second aspects, the crystal produced is one capable of diffracting X-rays or is one that is useful for X-ray analysis. As is explained below, the degree or resolution to which a crystal diffracts may be determined by the crystallization conditions. Preferably the crystal lattice produced by the first and second proteins is capable of diffracting x-rays to a resolution of at least 6Å, 5Å, more preferably at least 4Å, 3.5Å, 3.25Å or 3Å. Still more preferably, the crystal lattice can diffract x-rays to a resolution of more than 2.75Å or 2.5Å.

The second protein can be expressed together with the first protein as a fusion protein, or alternatively the second protein can be expressed

separately and positioned in the available space of the first protein by means of non-covalent interactions. The specificity and affinity necessary for this binding may be achieved by fusing protein tags or domains having such affinity for each other to the first and second protein, respectively. When the expression of a fusion protein is desired, the recombinant vector of the invention is adapted to allow for the insertion of at least one further nucleotide sequence, in particular a sequence encoding the second protein to be accommodated, upon its crystallization, in the said available space in the lattice of the first protein. It will be appreciated that the location of the nucleotide coding sequence encoding the second protein may be in one or more positions relative to the sequence encoding the first protein. In other words, the sequences encoding the first and second proteins may be in any order which provides for the second protein to be accommodated, upon its crystallisation, in the crystal lattice of the first protein. Hence, the sequences may be consecutive in any order, either contiguous or separated by a further sequence, or they may be non-consecutive, for example, the sequence of one protein may be inserted into the coding sequence of the other protein. Where the sequence of one protein is inserted into the coding sequence of the other protein, it is preferred if this is done such that the reading frame of the "other protein" is not changed.

The recombinant vector according to the invention comprises a promoter sequence operably linked to the structural gene(s) and is capable of mediating the expression of the said first protein or the said fusion protein. The term "operably linked" as used herein means that the promoter is functionally linked to a structural gene in the proper position to express the structural gene under control of the promoter.

The skilled person will be able to determine which proteins are suitable for use as the said "first protein" based on the crystal lattice structure of said

protein, the size of available cavities in this structure and the positions of sites available for attaching fusion partners. In determining whether a protein is suitable for use as the said "first protein", it will be appreciated that the size of available cavities or space in the crystal structure of the first protein when crystallized in the absence of a second protein may not be strictly limiting in practice. The first protein may assist in crystallization of the second protein even if the crystal space generated by the first protein and which is available to accommodate the second protein is not identical in the presence and absence of the second protein. This flexibility is sufficient so as to allow the first protein to modify its space group when crystallizing. Thus, the space group of the crystals of the first protein alone may differ from that of crystals of the first and second proteins when together.

Hence, it will be appreciated that the crystal lattice of the first protein useful in the present invention, when crystallised in the absence of a second protein, may have spaces or gaps which are solvent filled and which are smaller than the size of the second protein to be crystallised. Preferably the gaps or spaces in the crystal lattice of the first protein are at least the same size as the second protein.

The first protein may be any protein which is suitable for accommodating the second protein in its crystal lattice. Hence, the first protein may be a soluble protein, including a soluble multisubunit protein, or it may be a membrane protein including a membrane-associated protein or an integral membrane protein. Where the first protein is a multisubunit protein it may have any number of subunits, including 2, 3, 4, 5 or 6 or more. Preferably, where the first protein is a multisubunit protein, it is not an antibody or antibody fragment such as an Fv molecule or Fab-like molecule. Where the first protein is an integral membrane protein, it may have one transmembrane domain, or two or three or more transmembrane domains

(ie, it may be a polytopic membrane protein). Preferably the first protein is an integral membrane protein, and more preferably it has one transmembrane domain. Still more preferably the first protein has 2 or 3 or 4 or 6 or 7 or 12 transmembrane domains.

The first protein may be any size which, when crystallised, is capable of the required accommodation of the second protein. Preferably the first protein is bigger than 10 amino acids in total, more preferably bigger than 25, 50, 75 or 100 amino acids in total. Still more preferably, the first protein is bigger than 150aa, 200aa, 250aa, 300aa, 400aa, 500aa, 600aa, 700aa, 800aa or 1000aa in total length. By "total length" we mean the total number of amino acids in the first protein, including all component subunits where the first protein is a multisubunit protein.

It will be appreciated that in order to crystallise the first and second proteins together in a crystal lattice of a quality suitable for x-ray diffraction, it may be necessary to conduct a screen for suitable and optimal conditions for crystallisation. Suitable screens are discussed in more detail below. Such screening is routine in the art of crystallisation. A way of determining if a crystal is capable of diffracting x-rays is to observe the diffraction pattern.

Of the first and the second protein useful in the present invention, it may be determined which protein constitutes the crystal lattice by determining which protein contributes the greater proportion of crystal contacts which are made and maintained within the lattice.

In preferred forms of the first and second aspects of the invention, the said first protein is *E. coli* cytochrome *bo*3 *or E. coli* fumarate reductase, or variants thereof. As shown in Fig. 3, the crystals of cytochrome *bo*3 provide space for additional proteins at the C-terminal end of cytochrome

*bo*3 subunit IV. The term "variants thereof", as used herein, is intended to mean e.g. polypeptides carrying modifications like substitutions, small deletions, insertions or inversions, which polypeptides nevertheless have substantially the functional properties of *E. coli* cytochrome *bo*3 *or E. coli* fumarate reductase. It should be emphasized that the term "functional properties", as used in this context, does not refer to biological activity, but rather to the structural capability to assist in crystallization of the second protein, for example to harbor a second protein in its "available space" in order to facilitate crystallization of the said second protein.

The term "available space" is not to be construed as referring solely to a "cavity" or gap" within the crystal lattice of a said first protein. Rather, the available space also comprises solvent channels in the said crystal lattice. For instance, in the *bo*3 oxidase crystal (cf. Fig. 12), big "gaps" are repeating, which gaps are not isolated but connected by solvent channels. In the *bo*3-Apo AI crystal (Fig. 12), Apo AI is not staying in one gap, rather it extends through multiple gaps connected by solvent channels.

The second protein may be any second protein which is capable of being accommodated in the crystal lattice of the first protein. In a preferred embodiment, the second protein is smaller than the first protein. By "smaller" we include the meaning that the second protein has a lower molecular weight than the first protein. A lower molecular weight is any mass which is less than that of the first protein. Preferably, the molecular weight of the second protein is at least 1kDa, 5kDa, 10kDa, 15kDa or 20kDa lower than that of the first protein. More preferably the molecular weight of the second protein is smaller than that of the first protein by at least about 25kDa, 35kDa, 45kDa or 55kDa or more. Similarly, it is preferred if the second protein is not bigger than 150kDa, more preferably

no bigger than 125kDa, 110kDa, 100kDa or 90kDa. The size of the second protein may be less than 80kDa or 70kDa.

Hence, it will be appreciated that the term "available space" is one which indicates the space in a crystal lattice of a first protein, which space is not occupied by the first protein. This space therefore may be occupied by a second protein or by solvent molecules and still be referred to as "available space".

Furthermore, the term "available space" as defined above is not to be construed as referring only to a fixed volume within the crystal lattice of a said first protein in the absence of the second protein. Instead, the available space is one which is flexible and may alter in size and/or shape according to the nature of the second protein. In other words, the crystal group produced by crystallization of a first protein compared with that produced by crystallization of a first and second protein may not be the same.

Hence, in one embodiment of the first and second aspects of the invention, the crystal space group of the first protein on its own (ie, when the first protein is crystallized in the absence of the second protein) may be different to that obtained by crystallization in the presence of, or fused to, the second protein.

It will be appreciated that the second protein need not be fused to the first protein to allow its crystallization, and the first and second proteins may be, for example, produced as a fusion, cleaved, then crystallized.

According to another preferred embodiment of the first two aspects of the invention, the first and second proteins are fused to each other. The fusion may be "direct" such that the amino acid sequences of the two proteins are

contiguous, or "indirect", where the amino acid sequences of the two proteins are joined via a linking polypeptide sequence or sequences.

In a further preferred embodiment of the first and second aspects of the invention, the first protein is a multisubunit protein. The subunits may be held together by covalent or non-covalent bonds. An example of a multisubunit protein is *E. coli* cytochrome *bo*3.

Preferably, the first protein is one which, when crystallized in the absence of a second protein, its crystal lattice has solvent filled gaps of a suitable size for accommodating a second protein within the crystal lattice, whether or not the original space group of the first protein is maintained by accommodating the second protein. Hence, it will be appreciated that the first protein may be a soluble protein.

When the "first protein" is *E. coli* cytochrome *bo*3, the nucleotide sequence to be included in the recombinant vector of the invention is preferably selected from

(a) the polypeptide coding regions of the nucleotide sequence shown as SEQ ID NO: 13;

(b) nucleotide sequences, which e.g. can be at least 90% or 95% homologous, with the nucleotide sequence shown as SEQ ID NO. 13 in the Sequence Listing, and which are capable of hybridizing, under stringent hybridization conditions, to a nucleotide sequence complementary with the polypeptide coding regions of the nucleotide sequence as defined in (a); and

(c) other nucleic acid sequences encoding the same amino acid sequences as those defined in (a) or (b). Numerous such nucleotide sequences may be designed due to the degeneracy of the genetic code.

The term "stringent hybridization conditions" is known in the art from standard protocols (e.g. Ausubel *et al., supra)* and could be understood as e.g. hybridization to filter-bound DNA in 0.5 M $NaHP0_4$, 7% sodium dodecyl sulfate (SDS), 1 mM EDTA at +65C, and washing in 0. 1xSSC / 0.1% SDS at +68C.

It should thus be understood that the nucleotide sequence coding for cytochrome *bo*3, or a suitable variant thereof, is not limited strictly to the sequence shown as SEQ ID NO: 13. Rather this sequence is represented in DNA molecules carrying modifications like substitutions, small deletions, insertions or inversions, which nevertheless encode polypeptides having substantially the functional properties of *E. coli* cytochrome *bo*3. As mentioned above, the term "functional properties", does not in this context refer to biological activity, but rather to the structural capability to harbor a second protein in its "available space" in order to facilitate crystallization of the said second protein.

The promoter sequence to be included in the recombinant vector may be the one naturally associated with the DNA sequence encoding the first protein, or of another origin. When the said first protein is *E. coli* cytochrome *bo*3, the said promoter sequence could essentially comprise the cytochrome *bo*3 promoter sequence shown as positions 203 through 803 in SEQ NO: 1. Alternatively, the said promoter could be an inducible promoter, such as the promoter pBAD (Invitrogen Corp., CA, USA) as shown in Example 7, below.

In a preferred form of the invention, the recombinant vector can further comprise a nucleotide sequence encoding a linker amino acid sequence, facilitating for the said first and second proteins to be expressed as a fusion

protein. The ideal linker should have the appropriate length and flexibility so as to allow the second protein to be positioned in the available space of the crystal lattice, it should not form hydrophobic interactions with lipophilic structures such as host cell membranes or the protein core, it should not affect, in a negative way, the expression, translocation and folding of the fusion protein, it should not inhibit the functions of the first or second proteins and should be stable in the host cell and during purification. In addition, the linker may comprise sequences useful for the detection and/or purification of the fusion protein by means of affinity methods, especially when useful antibodies are unavailable. The said linker amino acid sequence can e.g. be a *Strep*-tag having an amino acid sequence shown as SEQ ID NO: 6 or a *Strep-HA*-tag, having an amino acid sequence shown as SEQ ID NO: 9. Preferably, the said linker amino acid sequence will be adapted to facilitate, upon expression of the said first and second proteins, for the said second protein to be positioned in the said available space in the lattice of the first protein. Thus, when the said first protein *is E. coli* cytochrome *bo*3, the said nucleotide sequence coding for a linker amino acid sequence can preferably be positioned at the 3'-end of the nucleotide sequence coding for *E. coli* cytochrome *bo*3 subunit IV.

As shown in Example 3, below, the recombinant vector according to the invention can in addition comprise a nucleotide sequence encoding a "protein Z" polypeptide having essentially an amino acid sequence shown as SEQ ID NO: 14. Since protein Z is a highly soluble and stable protein domain, its presence may facilitate the expression, solubilization, purification and crystallisation of the fusion protein.

The recombinant vector according to the invention may in addition comprise a nucleotide sequence encoding an affinity tag, e.g. a His-tag, useful for detection and or purification of the expressed protein(s). An

affinity tag is a protein sequence that provides a defined, strong, and highly specific non-covalent binding interaction to a ligand or another protein sequence or domain. The presence of the affinity tag in the expressed protein allows the detection and/or purification of said protein on the basis of a reversible interaction between said affinity tag and its specific ligand, said specific ligand being attached to an easily detected chemical entity e.g. a fluorophore or an enzyme) or chromatographic matrix, respectively. When the first protein *is E. coli* cytochrome *bo*3, the affinity tag could e.g. be attached to the nucleotide sequence encoding subunit II of *E. coli* cytochrome *bo*3 (cf positions 1746 through 1772 in SEQ ID NO: 1).

In a further important aspect of the invention, the recombinant vector defined above further comprises a nucleotide sequence encoding the said "second" protein. Preferably, the second protein is one as described above. When cytochrome *bo*3 is the first protein, it is estimated that the "cavity" in the cytochrome *bo*3 crystal (cf. Fig. 3) can harbor a protein having a molecular mass up to approximately 100 kDa.

As explained above, the presence of a second protein in a crystal of the first protein may alter the space group of the first protein from that which is formed when the first protein is crystallized in the absence of the second protein. Hence, it will be appreciated that the "cavity" in the cytochrome *bo*3 crystal may in fact be capable of harbouring a protein larger than 100kDa. Hence, due to the flexibility of the crystal space group, the predicted size of the "cavity" is not to be considered limiting in the choice of second protein.

Consequently, at least when cytochrome *bo*3 is the first protein, the said second protein preferably has a molecular mass below 100 kDa, such as below 75 kDa, below 60 kDa or below 50 kDa. The skilled person will be

able to determine the possible size of the second protein, depending on the crystal lattice structure and available positions for the attachment of fusion partners in the first protein to be used. Furthermore, the second protein must be expressed in the system used in a correctly folded form and be able to translocate within the host cell in a manner consistent with the intended subcellular location and orientation of the fusion protein. In addition, the function of the second protein should be maintained when expressed in the system used, so as to allow a functional assay to be performed, demonstrating that the second protein is in its native or native-like form.

In a preferred embodiment of the invention, the second protein is a membrane protein, and more preferably, it is an integral membrane protein. Such an integral membrane protein may have any number of transmembrane domains, ranging from one to twelve or more.

Preferably, the second protein has a lower molecular weight compared to the first protein.

Included in the invention is also a cultured host cell, e.g. an *E.coli* cell, harboring a recombinant vector according, to the invention, in particular a recombinant vector comprising a nucleotide sequence encoding for a said second protein.

A further aspect of the invention is a process for the production of a fusion protein comprising culturing a host cell as defined above, under conditions whereby the said fusion protein is produced, and recovering the said fusion protein. A fusion protein obtained, or obtainable by this process is included in the invention.

In the case that the recombinant protein is a multisubunit complex kept together by non-covalent forces the nucleic acid sequences encoding the individual subunits of said complex may be introduced into the host organism by use of more than one vector, each vector encoding one or more of these subunits.

A further aspect of the invention provides a fusion protein comprising (i) a first protein which is a membrane protein or multisubunit protein and which, when crystallized with a second protein, is capable of accommodating the second protein in the crystal lattice and (ii) a second protein to be located, when crystallized, in the crystal lattice of the first protein wherein the resulting crystal lattice is capable of diffracting x-rays.

In yet another aspect, the invention provides a fusion protein comprising (i) a first protein which is a membrane protein or multisubunit protein and which upon crystallization yields crystals having available space in the lattice, so as to allow for the ordered packing of a second protein into the said available space; and (ii) a second protein to be accommodated, upon crystallization, in the said available space wherein the resulting crystal lattice is capable of diffracting x-rays.

When the said first protein is *E. coli* cytochrome *bo*3, the said second protein is preferably attached to subunit IV of *E. coli* cytochrome *bo*3.

In a preferred embodiment, either or both of the proteins comprised in the fusion proteins of the invention are membrane proteins. In other words, either the first or second proteins, or both of them, are membrane proteins.

By "membrane protein" we include membrane associated proteins, membrane inserted proteins (such as those which possess a hydrophobic

domain which is resident within a membrane but may not completely span the membrane bilayer) and integral membrane proteins where the protein possesses at least one transmembrane domain which spans the membrane bilayer, such as single spanning integral membrane proteins and polytopic membrane proteins. Preferably, the membrane protein is an integral membrane protein.

More preferably, the first protein is one as defined above in relation to the first and second aspects of the invention.

In a further aspect, the invention provides a method for crystallization of a protein, comprising (i) obtaining a fusion protein comprising (I) a first protein, which is a membrane protein or multisubunit protein and which upon crystallization yields crystals having available space in the lattice, so as to facilitate crystallization of a second protein; and (II) the said (second) protein to be crystallized; and (ii) crystallizing the said fusion protein wherein the resulting crystal is capable of diffracting x-rays.

The said fusion protein could preferably be obtained by the processes described above, in particular by expression of the recombinant vectors according to the invention.

Preferably the first protein is a membrane protein, and more preferably it is an integral membrane protein.

A still further aspect of the invention provides a method for crystallization of a protein, comprising (i) obtaining a first protein which is an integral membrane protein and which upon crystallization yields crystals having available space in the lattice, so as to facilitate crystallization of a second protein; and (ii) obtaining the second protein to be crystallized; and (iii)

crystallising both the said proteins together wherein the resulting crystal is capable of diffracting x-rays.

The first and second proteins of this aspect could be produced as a fusion protein, preferably obtained by the processes described above, in particular by expression of the recombinant vectors according to the invention.

In an embodiment of this aspect of the invention, the proteins could be obtained by expressing the first and second proteins in two separate expression systems and fusing them after purification, either by simply mixing the two purified protein samples or by soaking that second protein into crystals of the first protein. It will be appreciated that in order to introduce the second protein into the solvent gap of the crystals of the first protein by soaking it is necessary to have a means of targeting the second protein to the precise location within the crystal lattice of the first protein, i.e. some form of protein-protein interaction. This can be achieved using high affinity domains engineered into suitable sites within the proteins, which when a suitable concentration of the second protein is added to the crystals of the first protein, allows the proteins to form a complex based on the affinity of the two domains.

Preferably, the first protein useful in the methods of crystallisation according to the present invention is as defined above in respect of the recombinant vectors of the invention.

In a preferred embodiment of the crystallisation methods of the invention, the second protein is as defined according to the first or second aspects of the invention.

The second protein may be a soluble protein, a membrane associated protein or an integral membrane protein. The method is particularly useful where the second protein is an integral membrane protein. Hence it is preferred if the second protein is an integral membrane protein.

In a preferred embodiment of the crystallisation method aspects of the invention, the method further comprises a step wherein at least two detergents are screened in the crystal growth conditions to identify which one optimises the growth and/or diffraction of the resulting crystals. It is known that detergent selection is important to obtain well diffracting crystals. In the case where cytochrome *bo*3 is the first protein, a wide range of detergents are tolerated, providing a broad choice of detergent in order to maximise the resolving ability of the crystals produced.

Suitable detergents for screening include all detergents of the C7-C9 range. Preferably, one of the detergents screened is octylglucoside.

Optimal growth of the crystals is that which gives a smaller number of crystals, preferably a single crystal, which are large. Preferably the resulting crystal is assessed by its diffraction pattern, with those crystals which produce a diffraction pattern being preferred to those which do not.

In a further preferred embodiment of these crystallisation method aspects, the pH of crystallisation is optimised for crystal growth. The pH screening may be performed between a range of pH 6-8. Typically, an initial screen to optimise pH of crystallisation may test pH values of 6, 6.5, 7, 7.5 and 8. Preferably, this embodiment comprises a further screen for an optimised pH wherein the optimal pH value identified by the initial pH screen is tested to a finer degree. For example, where pH 6.5 is identified as optimal in the

first screen, a subsequent pH screen may test pH values of 6.3, 6.4, 6.5, 6.6, 6.7 and 6.8.

According to a yet further preferred embodiment, the optimisation of crystallisation pH is performed in addition to a screen to identify a detergent which optimises the growth of well-diffracting crystals.

When the second protein is a hydrophobic protein, such as an integral membrane protein, an altered pH and a higher PEG concentration (relative to the conditions employed for crystallization of the first protein alone) may be necessary.

PEG acts as a precipitant in crystallisation, and acts to alter the protein-solvent or protein-protein contacts so that the protein molecules precipitate out of solution, preferably as ordered crystals. Other precipitants are known to be useful in crystallisation, including, for example, ammonium sulphate and 2-methyl-2,4-pentanediol (MPD). Alternative precipitants are give in Bergfors (1999) in *Protein Crystallization* Ed. Terese Bergfors, International University Line pp 41-50.

It will be appreciated that the crystal space group of the first protein alone (ie, in the absence of the second protein) may be different to that obtained by crystallization of the first protein with the second protein.

An additional aspect of the invention provides a method of obtaining structural data on a protein of interest comprising the steps of

(i) obtaining the protein of interest;

(ii) crystallising said protein in the crystal lattice of another protein, which crystal lattice is able to accommodate the protein of interest; and

(iii) obtaining x-ray diffraction data from the crystal produced in step (ii).

The crystallisation method used in step (ii) may be any suitable method. Preferably it is a method according to the present invention as described above.

The protein of interest may be obtained by any convenient method. Advantageously, the protein of interest is obtained by expressing a recombinant vector according to the present invention or by culturing a cell according to the invention.

The protein of interest may be any protein for which structural data is desired. Preferably, the protein of interest is one according to the definition of the "second protein" given above. More preferably, the protein of interest is an integral membrane protein.

Typically, the x-ray diffraction data is obtained to a level of resolution which can yield structural information. This resolution may vary according to the detail of structural information required. Preferably, the resolution is at least 6Å, more preferably at least 5Å, still more preferably at least 4, 3.5, 3.2, or 3, or 2.5Å.

The present invention further provides a use of a recombinant vector, or a cell, according to the invention, in a method of obtaining structural data on a protein of interest according to the invention.

In yet a further aspect, the invention provides a process for the production of a recombinant vector according to the invention comprising

(i) obtaining a recombinant vector comprising (I) a nucleotide sequence encoding a first protein which is a membrane protein or multisubunit protein and which, when crystallized with a second protein, is capable of accommodating the second protein in the crystal lattice and (II) a promoter operably linked to the said nucleotide sequence; and

(ii) introducing, into the said vector, nucleotide sequences facilitating the insertion of further nucleotide sequences wherein the resulting crystal would be capable of diffracting x-rays.

The nucleotide sequences may be sequences including a restriction endonuclease cleavage site.

An additional aspect of the invention provides a process for the production of a recombinant vector according to the invention, comprising

(i) obtaining a recombinant vector comprising (I) a nucleotide sequence encoding a first protein which is a membrane protein or multisubunit protein and which upon crystallization yields crystals having available space in the lattice, so as to allow for the ordered packing of a second protein into the said available space, and (II) a promoter operably linked to the said nucleotide sequence; and

(ii) introducing into the said vector, nucleotide sequences facilitating the insertion of further nucleotide sequences wherein the resulting crystal would be capable of diffracting x-rays.

The further nucleotide sequences are preferably sequences encoding at least one further protein which is the protein to be crystallized.

The said recombinant vector obtained in step (i) could e.g. be the vector designated pMB908 (disclosed as "pJRhisA" by Rumbley *et al.* (1997)

Biochimica et Biophysica Acta 1340, 131-142 which comprises the nucleotide sequence shown as SEQ ID NO: 1.

The nucleotide sequences may be sequences including a restriction endonuclease cleavage site.

EXPERIMENTAL METHODS

The modified vector constructs were generated using standard methods, such as molecular cloning methods, PCR, restriction analysis, DNA preparative methods, etc. In this context, the term "standard methods" is to be understood as referring to protocols and procedures found in an ordinary laboratory manual such as: Current Protocols in Molecular Biology, editors F. Ausubel *et al.*, John Wiley and Sons, Inc. 1994, or Sambrook, J., Fritsch, E.F. and Maniatis, T., Molecular Cloning: A laboratory manual, 2nd Ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY 1989.

Large-scale production of the cytochrome *bo*3 fusion protein was performed in *E. coli,* using a 10L-fermentor. *E. coli* cells were grown in LB supplemented with 1 g $KH2PO_4$, 14.6 g $K2HPO_4$, 20 ml Na-lactate, 0.5 ml 1 M $MgSO_4$, 0.00214 g $CuSO_4$, 0.0102g vitamin B1, 0.0098 g Nicotinic Acid and 0.05 g ampicillin/L over a period of 8 hrs (Georgiou *et al.* (1988) Biochim. Biophys. Acta 933, 179-183).

EXAMPLES OF THE INVENTION

EXAMPLE 1: Crystallization and structural determination of cytochrome *bo*3

The structure of cytochrome *bo*3 from *Escherichia coli* was determined. Plasmids encoding a carboxy-terminus histidine-tag on subunit II of cytochrome *bo*3 ubiquinol oxidase were cloned into an *E. coli* strain, G0105, lacking terminal oxidases as described in Kaysser *et al.* (1995) Biochemistry 34, 13491-13501. Purified protein was crystallized using polyethylene glycol 1500. Data collection from the crystal was performed at ID14/EH3 of the ESRF. Image data was processed up to 3.5Å resolution (Fig.1) with an R_{merge} value of 10.8% (for F>1.0 (F)). The crystals belong to the space group C2221 with unit cell dimensions of a=92.1 Å, b=372.5 Å and c=232.7 Å. The asymmetric unit contained two molecules of ubiquinol oxidase. The *bo*3 protein has a molecular weight of 144 kDa and occupies about 41% of the volume of the unit cell.

EXAMPLE 2: Expression vector constructs

A multiple cloning site *(Not*I, *Sac*I, *Mlu*I *and Xba*I*)* was added to the 5'-end of subunit IV of the cytochrome bo3 scene. A plasmid designated pMB908 (SEQ ID NO:1), which comprises a cytochrome *bo*3 construct with a His9 tag at the C-terminus of subunit II was used as starting material. The plasmid pMB908 is identical to the pJRhisA plasmid described by Rumbley *et al.* (1997) Biochimica et Biophysica Acta 1340, 131-142. The addition of unique restriction sites to the carboxy-terminus to subunit IV in pMB908 was performed by the polymerase chain reaction (PCR) method of splicing by overlap extension. The method entails the use of four different primers, two encompassing the entire sequence to be changed, one on either end of the new construct (the 5'-primer is set forth as SEQ ID NO: 2 and the 3'-primer as SEQ ID NO: 3), and two centrally positioned primers (The 5'-primer is set forth as SEQ ID NO: 4 and the 3'-primer as SEQ ID NO: 5) containing the sequence for the restriction sites *(Not*I, *Sac*I, *Mlu*I, *Xba*I*)* giving rise to two fragments containing overlapping sequences. The PCR

reactions are carried out in two steps to yield a 2.4 kb fragment of the cytochrome *bo*3 gene which was flanked by two unique restrictions sites *(Nsi*I *and Sph*I*)*. This fragment was sequenced and then heated into the cytochrome *bo*3 gene generating the construct MB930 (Fig. 4).

Two sets of linkers were generated to act as bridge sequences between subunit IV and the foreign fusion protein. A short olioonucleotide linker, coding for the *Strep*-tag, a nonapeptide having the sequence AWRHPQFGG (SEQ ID NO: 6) was formed by annealing two single-stranded, synthetic oligonucleotides (The forward-strand oligonucleotide is set forth as SEQ ID NO: 7 and the reverse strand oligonucleotide as SEQ ID NO: 8), coding for this sequence and containing a 5'-*Not*I site and 3'-*Mlu*I site. The *Strep*-tag is an amino acid sequence which was identified using phage display based on its affinity for streptavidin (Schmidt & Skerra (1993) Protein Engineering 6, 109-122). A second linker (Strep-HA-tag; AWRHPQFGGYPYDVPDYA) (SEQ ID NO: 9) coding for both the *Strep*-tag and the hemaglutinin (HA)-tag YPYDVPDYA (SEQ ID NO:10) (Kast *et al.* (1996) *J. Biol. Chem.* 271(16), 9240-9248), was made by annealing, linear, single-stranded, synthetic oligonucleotides (The forward-strand oligonucleotide is set forth as SEQ ID NO: 11 and the reverse-strand oligonucleotide as SEQ ID NO: 12). The oligonucleotide cassettes were generated by mixing 10nmol of 5'- and 3'-oligonucleotide with 30 1 annealing buffer (500 MM NaCl, 100 nM Tris-HCl, pH 7.4, and 100 nM $MgCl_2$) in a total volume of 300 1. The samples were boiled for two minutes and then allowed to cool to approximately +30C prior to storage at -20C. The linkers were cloned into the *Not*I / *Mlu*I sites of the construct pMB930, yielding constructs pMB946 and pMB947 coding for the *Strep* and *Strep-HA* linker sequences, respectively.

The expression of all the modified cytochrome *bo*3 constructs was assessed in two ways. Firstly, each construct was transformed into the *E. coli* strain G0105 Kaysser et al. (1995) Biochemistry 34:13491-13501. These cells lack an endogenous oxidase activity and will only grow under aerobic conditions after the introduction of a functional oxidase. All the vector constructs produced GO105 cell colonies, indicating that the additions to the sequence had not altered the function of the cytochrome *bo*3. Secondly, a His-tag was present at the C-terminal end of subunit II of the cytochrome *bo*3, and in pMB947 there was a HA-tag at the C-terminal end of subunit IV. It was thus possible to assess the expression of the constructs using Western blot analysis (Fig. 5) with antibodies directed against these specific sequences.

EXAMPLE 3: Cytochrome bo3 - Protein Z fusion protein

The polypeptide designated "Protein Z" or "Domain Z" (Nilsson et al. (1987) Prot. Eng., 107-13; SEQ ID NO: 14) is a modified analogue of the IgG-binding domain B of *Staphylococcus aureus* protein A (SPA). Protein Z (6.6 kDa) has been extensively used as an affinity tag for reviews see Nilsson *et al.* (1992) *Curr. Opin. Struct. Biol.* 2: 5 69575; and LaVallie McCoy (1995) Curr. Opin. Biotech 6: 501-506. The structure of SPA domain B has been resolved to 2.8Å (Deisenhofer (1981) Biochemistry 20, 236170).

Protein Z was cloned into the *Not*I and Mlu*I* sites of the cytochrome *bo*3 fusion vector using standard methods yielding pMB1048. Expression of the cytochrome *bo*3 fusion construct was assessed by Western blot analysis and cells expressing the fusion protein were grown in a fermentor (OD_{550}=3.0), harvested and stored at -80C. Membranes were purified by the following method: Cells from a 1OL culture were taken to a final

volume of 1L in lysozyme treatment buffer (200 mM Tris-HCl, pH 8.8, 20 mM EDTA, pH 8.0, and 500 mM sucrose), lysozyme was added to a final concentration of 0.1 % and the cells stirred for 30 min. The cells were pelleted by centrifugation at 8,000 rpm for 20 min. The pellets were resuspended in approx. 750 ml cell disruption buffer (5 mM EDTA, pH 8.0, 10 M PMSF, 10 mM $MgCl_2$ and several crystals of DNAse I) and stirred on ice for 15 min. The solution was sonicated on burst mode for 2 x 3 min. Unbroken cells were separated by centrifugation at 6000 rpm for 20 min and the supernatant centrifuged 45,000 rpm for 1 h. The membrane pellets were resuspended in a minimal volume of buffer (20 mM Tris-HCI, pH 7.5, 300 mM NaCl and 2.5 mM imidazole). At this point it was possible to freeze the membrane pellets at -80C prior to solubilization and purification.

The membranes, were solubilized in 1 % dodecylmaltoside for 1h, +4C. Solubilized protein was harvested after ultracentrifugation at 100,000 x g for 1 h and applied to a 65ml bed volume Ni-NTA column (Qiagen) equilibrated with 20 mM Tris-HCl, pH 7.5, 300 mM NaCl, 5 mM imidazole and 0.03% dodecylmaltoside. The column was washed with three bed volumes of the equilibration buffer to remove any non-specific binding proteins and then the sample was eluted with a linear 5-150 mM imidazole gradient. The resulting chromatograms showed two clear protein peaks, which were termed "low" and "high" imidazole based on the elution concentration.

The fractions from the two peaks were pooled separately. The buffer was then exchainged for 20 mM Tris-HCl, pH 7.5 containing 0.03% dodecyl maltoside using an Amicon stirred ultrafiltration cell (100 kDa cutoff filter) (Amicon, MA, USA). The two pools were then applied separately to an anion exchange column, MonoQ 10/10 (Amersham Pharmacia

Biotech, Sweden) in the presence of 1 mM potassium ferricyanide after equilibrating the column with 20 mM Tris-HCl, pH 7.5 containing 1% octylglucoside. The column was washed slowly with 6 bed volumes of equilibration buffer and then eluted with a 0-600 mM NaCl gradient. Fractions containing the fusion protein were pooled on the basis of spectrophotometric readings. Using a Centricon ultrafiltration cell (100 kDa cutoff) (Millipore, MA, USA), the buffer was exchanged for 20 mM Tris-HCl, pH 7.5 containing, 1% octylglucoside, and the sample was concentrated to 20 mg protein/ml.

Crystals for this fusion protein (Fig. 6) were obtained for "low" and "high" imidazole protein, using the hanging drop vapor diffusion technique. The protein solution contained 20 mM Tris-HCl, pH 7.5, and 1% octylglucoside. A reservoir solution of 9-10% PEG 1500, 100 mM NaCl, 100 mM $MgCl_2$ and 5% ethanol was used. The protein solution was mixed in a 1: 1 ratio with the reservoir solution and left to equilibrate at +4C. These data demonstrate that the site at the C-terminal end of subunit IV can be used to accommodate foreign proteins as fusion partners without inhibiting crystal formation. Crystals of cytochrome *bo3* + Z, were obtained under similar conditions to the native cytochrome *bo3* (Abramson *et al*, 2000), although the crystals of the fusion protein grew over a wider pH range (6-8) compared to the native protein (pH 7-7.5). In addition it was possible to grow crystals using protein from both the "low" and "high" imidazole peaks, compared to the native protein which only yielded reproducible crystals from the "low" imidazole peak. The crystals of the fusion protein from both "low" and "high" imidazole peaks grow as square plates, a more regular shape compared to the rod-like crystals of the native cytochrome *bo3*. Data was collected from a cytochrome *bo3* + protein Z fusion protein crystal which diffracted X-ray up to 6 Å and had a space group $P2_1$ with the

cell dimensions *a*=93.4 Å, *b*= 328.7 Å, *c*= 131.1 Å and *=92.1 with four molecules in the assymetric unit. These data are in contrast with the wild-type protein crystals which have a space group $C222_1$ with the cell dimensions *a*=91.3 Å, *b*= 370.3 Å, *c*= 232.4 Å (see example 1).

EXAMPLE 4: Cytochrome *bo*3 - GPCR fusion proteins

Two G-protein coupled receptors (GPCRs); the human muscarinic 1 (M1) receptor of 51 kDa (Allard *et al.* (1987) *Nucleic Acids Res* 15: 10604) and the human canabinoid 2 (CB2) receptor of 40 kDa (Munro *et al.* (1993) *Nature* 365, 61-65) were cloned into the *Mlu*I and *Xba*I sites of the cytochrome *bo*3 fusion vectors according to standard methods. The receptors expressed as fusion partners with subunit IV of the cytochrome *bo*3 (Fig. 7).

EXAMPLE 5:
Cytochrome *bo*3 - Leader peptidase fusion protein and cytochrome *bo*3 ProW fusion protein.

A fusion construct of cytochrome *bo*3 + *E. coli* leader peptidase (Whitchurch & Mattick (1994) *Gene* 150: 9-15) (no tag sequence in vector) was obtained as described above. The biological role of the *E. coli* leader peptidase (36 kDa) is to remove amino-terminal leader peptides from exported proteins after they have crossed the plasma membrane. The enzyme of 323 amino acid residues spans the membrane twice, with its large carboxy-terminal domain protruding into the periplasm (for a review, see Dalbey (1991) *Mol. Microbiol.* 5, 2855-2860).

A fusion construct of cytochrome *bo*3 + *E coli* ProW (Growshiker (1989) J. Bacteriol. 171: 1923-193 1) *(+ Strep*-tag) was obtained as described above. ProW is an *E. coli* inner membrane protein of 38 kDa that consists of a 100-residue-long periplasmic N-terminal tail followed by seven closely spaced transmembrane segments (Cristobal *et al.* (1999) *J. Biol. Chem.* 274, 20068-20070). It is part of the ProU system, a member of the ATP-binding cassette (ABC) superfamily of transporters (Lucht & Bremer (1994) *FEMS Microbiology Letters* 14: 3-20). Both fusion constructs were shown to express (Fig. 8).

EXAMPLE 6: Cytochrome bo3 - Apo AI fusion protein

Apolipoprotein AI (Apo AI, Sharpe *et al.* (1984) *Nucleic Acids Res.* 12: 3917-3932) is the major protein component (28 kDa) of the serum high-density lipoprotein (HDL) particles (for a review, see Hargrove *et al.* (1999) *J. Mol. Endocrinol.* 22, 103-111). The structure of truncated human Apo AI has been determined at 3Å resolution (Borhani *et al.* (1999) *Acta Cryst. D* 55: 12291-12296).

The following fusion constructs of cytochrome *bo*3 and Apo AI were generated:

pMB1241 Cytochrome *bo*3 + Apo AI (no tag)
pMB1242 Cytochrome *bo*3 + *Strep*-tag + Apo AI
pMB1243 Cytochrome *bo*3 + *Strep-HA*-tag + Apo AI
pMB1244 Cytochrome *bo*3 + Protein Z + Apo AI

Interestingly, the expression of these fusion proteins was different from the earlier expressed proteins (Fig. 9). No expression of cytochrome *bo*3 +

Strep-HA-tag + Apo AI (pMB1243) could be detected. The cytochrome *bo*3 + Protein Z + Apo AI (pMB1244) exhibited the highest level of expression, although it underwent a certain amount of proteolytic degradation. Cytochrome *bo*3 + Apo AI (no tag) exhibited detectable levels of expression (pMB1241) although it too degrades (50% loss). The cytochrome *bo*3 + *Strep*-tag + Apo AI (pMB1242) construct expresses at relatively high levels and does not appear to be degraded. In addition, the fusion proteins only appear to express satisfactorily to an OD600 of approximately 1, after which the whole cytochrome *bo*3 Apo AI complex appears to be broken down. These results show the individuality of proteins expressed in this system and highlight the need to characterize the growth and production of each fusion protein. Optimizing the construct as well as the host system and cultivation conditions with regard to the properties of the target protein may be performed by a person skilled in the art as shown in the current example.

It was possible to obtain crystals of pMB1242, cytochrome bo3 + apo A-I. These had an elongated hexagonal plate form. In addition, these crystals were also thicker and had sharper edges than the native crystals, indicating ordered packing within the crystal lattice. These crystals diffracted to 5 Å and belong to the space group C2 with the cell dimensions a=93.4 Å, b= 328.7 Å, c= 131.1 Å and *=92.1 and two molecules per assymmetric unit. Crystals of pMB1246, cytochrome bo3 + ProW, were obtained under slightly different conditions; PEG 1500 concentration was 18% and pH 6.5. These crystals diffracted to 6 Å and belong to the space group C2 with the cell dimensions a=93.4 Å, b= 328.7 Å, c= 131.1 Å and *=92.1 and two molecules per assymmetric unit. The wide-belt like loop of the apolipoprotein circles the whole cytochrome bo3 molecule and forms protein-protein contacts within the lattice. It is possible that given the function of apo A-I *in vivo*, it is binding to the detergent micelle which

surrounds the hydrophobic portions of cytochrome bo3. The apolipoprotein A-I appears to be accommodated within multiple gaps in the crystal lattice which are connected by solvent channels.

EXAMPLE 7: Expression of cytochrome *bo*3 under the control of an inducible promoter.

The constitutive cytochrome *bo*3 promoter was replaced with an inducible pBAD promoter cloned using the pBADHis vector (Invitrogen Corp., CA, USA) as a template. The pBAD Expression System is based on the *araBAD* operon, which controls the arabinose metabolic pathway in *E. coli.* This construct was expressed in an alternative cell line, GL1O1 (Rumbley *et al.* (1997) *Biochem. Biophys. Acta* 1340: 131-142). These cells express a form of oxidase and thus will grow under aerobic conditions in the absence of *bo*3. The expression of cytochrome *bo*3 was induced with increasing concentrations of arabinose, once the cells reached an OD600 of 0.5. Maximum expression of the cytochrome *bo*3 construct was observed at 0.2% arabinose with higher concentrations causing no further significant increase in expression.

The expression of cytochrome *bo*3 under control of the inducible promoter (pMB1127) was higher than that under control of the constitutive promoter (pMB908) (Fig. 10). Time course studies showed that the maximum detectable expression took place within 3-4 hours post induction (Fig. 11).

pBAD expression vectors were (generated for cytochrome *bo*3 with the MCS, the *Strep*tag, the *Strep-HA*-tag and protein Z yielding plasmids pMB1271, pMB1128, pMB1272, pMB1270 respectively. Nucleotide sequences coding for polypeptides such as e.g. Apo AI or the CB2 receptor

can be cloned into these vectors, and the expression of such polypeptides under control of the inducible promoter can be determined.

EXAMPLE 8: Use of fumarate reductase as a scaffold molecule

The *E. coli* respiratory enzyme fumarate reductase (FRD) is a four-subunit protein with a molecular mass of 121 kDa, which catalyzes fumarate reduction to succinate using membrane-bound menaquinol in anaerobic respiration (Kroger (1978) *Biochim. Biophys. Acta* 505, 129-145; Cole *et al.* (1985) *Biochim. Biophys. Acta* 811, 381-403).

Recently, the structure of FRD has been solved to 3.3 Å (Iverson *et al.* (1999) Science 284, 1961-1966) and subsequently to 2.8 Å (Tina Iverson, personal communication). In similarity with cytochrome *bo*3, the crystal lattice of FRD incorporates a gap, which could be exploited for crystallization of heterologous proteins. Two subunits of FRD, FrdC and FrdD, have three transmembrane helices and it would be possible to make fusion proteins at the C-terminal of these subunits using similar constructs as those described for cytochrome *bo*3. The expression level of FRD is very high under anaerobic conditions and purification is facilitated by the fact that the protein is selectively solubilized by the nonionic detergent Thesit (Maklashina (1998) *J. Bacteriol.* 180, 5989-5896), also called Polidocanol. The same restriction sites at the 3'-ends of FrdC and FrdD as for cytochrome *bo*3 subunit IV are generated, to allow simple transfer of nucleic acid sequences encoding the target proteins between scaffold molecules.

CLAIMS

1. A recombinant vector comprising, (i) a promoter sequence and (ii) a nucleotide sequence encoding a first protein which is a membrane protein or multisubunit protein and which, when crystallized with a second protein, is capable of accommodating the second protein in the crystal lattice; said recombinant vector further allowing for the insertion of a further nucleotide sequence encoding a second protein to be located, when crystallized, in the crystal lattice of the first protein wherein the resulting crystal lattice is capable of diffracting x-rays.

2. A recombinant vector comprising, (i) a promoter sequence and (ii) a nucleotide sequence encoding a first protein which is a membrane protein or multisubunit protein and which upon crystallization yields crystals having available space in the lattice, so as to allow for the ordered packing of a second protein into the said available space; said recombinant vector further allowing for the insertion of a further nucleotide sequence encoding a second protein to be accommodated, upon its crystallization, in the said available space in the lattice of the first protein wherein the resulting crystal lattice is capable of diffracting x-rays.

3. A recombinant vector according to Claim 1 or 2 wherein the x-ray diffraction is to a resolution of at least 5Å.

4. A recombinant vector according to Claim 3 wherein the diffraction resolution is at least 4Å.

5. A recombinant vector according to any one of Claims 1 to 4 wherein the first protein is a fusion partner of the second protein.

6. A recombinant vector according to Claims 1 to 5 wherein the crystal space group of the first protein when crystallised alone may be different to that obtained by crystallisation with the second protein.

7. A recombinant vector according to any one of Claims 1 to 6 wherein the said nucleotide sequence encoding a first protein is a sequence encoding a multisubunit protein.

8. The recombinant vector according to any one of Claims 1 to 7 wherein the said nucleotide sequence encoding a first protein is a sequence encoding a membrane protein.

9. The recombinant vector according to any one of Claims 1 to 8 wherein the said nucleotide sequence encoding a first protein is a sequence encoding an integral membrane protein.

10. The recombinant vector according to Claim 9 wherein the integral membrane protein has one transmembrane domain.

11. The recombinant vector according to any one of Claims 1 to 10 wherein the size of the first protein encoded by the nucleotide sequence is more than 10 amino acids in total.

12. The recombinant vector according to any one of Claims 1 to 10 wherein the said nucleotide sequence encoding a first protein is a sequence encoding *E. coli* cytochrome *bo*3 or *E. coli* fumarate reductase, or variants thereof.

13. The recombinant vector according to Claim 12 wherein the said nucleotide sequence encoding a first protein is a sequence encoding *E. coli* cytochrome *bo*3 or a variant thereof.

14. The recombinant vector according to Claim 13 wherein the said nucleotide sequence encoding *E. coli* cytochrome *bo*3 *is* selected from

(a) the polypeptide coding regions of the nucleotide sequence shown as SEQ ID NO: 13;

(b) nucleotide sequences capable of hybridizing under stringent hybridization conditions, to a nucleotide sequence complementary with the polypeptide coding regions of the nucleotide sequence as defined in (a), and

(c) nucleic acid sequences which are degenerate as a result of the genetic code to a nucleotide sequence as defined in (a) or (b).

15. The recombinant vector according to Claim 13 or 14 wherein the said promoter sequence essentially comprises the cytochrome *bo*3 promoter sequence shown as positions 203 through 803 in SEQ NO: 1.

16. The recombinant vector according to any one of Claims 1 to 15 wherein the said promoter is an inducible promoter.

17. The recombinant vector according to any one of Claims 1 to 16, further comprising a nucleotide sequence encoding a linker amino acid sequence facilitating for the said first and second proteins to be expressed as a fusion protein.

18. The recombinant vector according to Claim 17 wherein the said linker amino acid sequence is adapted to facilitate, upon expression of the said

first and second proteins, for the said second protein to be positioned in the said available space in the crystal lattice of the first protein.

19. The recombinant vector according to Claim 17 or 18 wherein said linker amino acid sequence is *a Strep*-tag having an amino acid sequence shown as SEQ ID NO: 6.

20. The recombinant vector according to Claim 17 or 18 wherein said linker amino acid sequence is a *Strep-HA*-tag having an amino acid sequence shown as SEQ ID NO: 9.

21. The recombinant vector according to any one of Claims 17 to 19 wherein the said nucleotide sequence coding for a linker amino acid sequence is positioned at the 3'-end of the nucleotide sequence coding for *E. coli* cytochrome *bo*3 subunit IV.

22. The recombinant vector according to any one of Claims 1 to 21 in addition comprising a nucleotide sequence encoding a polypeptide having essentially an amino acid sequence shown as SEQ ID NO: 14.

23. The recombinant vector according, to any one of Claims 1 to 22, in addition comprising a nucleotide sequence encoding an affinity tag.

24. The recombinant vector according to Claim 23, wherein the said affinity tag is a His-tag.

25. The recombinant vector according to Claim 23 or 24, wherein the said first protein *is E. coli* cytochrome *bo*3 and wherein a nucleotide sequence encoding an affinity tag is attached to the nucleotide sequence encoding *E. coli* cytochrome *bo*3 subunit II.

26. The recombinant vector according to any one of Claims 1 to 25, further comprising a nucleotide sequence encoding the said second protein.

27. The recombinant vector according to Claim 26, wherein the said second protein has a molecular mass below 100 kDa.

28. The recombinant vector according to Claim 26 or 27 wherein the second protein has a lower molecular weight than the first protein.

29. The recombinant vector according to any one of Claims 22 to 28 wherein the said second protein is a membrane protein.

30. A cultured host cell harbouring a recombinant vector as defined in any one of Claims 26 to 29.

31. The host cell according to Claim 30 which is an *E. coli* cell.

32. A process for the production of a fusion protein which comprises culturing a host cell as defined in Claim 30 or 31 under conditions whereby the said fusion protein is produced, and recovering the said fusion protein.

33. A fusion protein obtained or obtainable by the process as defined in Claim 32.

34. A fusion protein comprising (i) a first protein which is a membrane protein or multisubunit protein and which upon crystallization yields crystals having available space in the lattice, so as to allow for the ordered packing of a second protein into the said available space; and

(ii) a second protein to be accommodated, upon crystallization, in the said available space wherein the resulting crystal is capable of diffracting x-rays.

35. A fusion protein comprising (i) a first protein which a first protein which is a membrane protein or multisubunit protein and which, when crystallized with a second protein, is capable of accommodating the second protein in the crystal lattice and (ii) a second protein to be located, when crystallized, in the crystal lattice of the first protein wherein the resulting crystal lattice is capable of diffracting x-rays.

36. A fusion protein according to Claim 34 or 35 wherein either or both of the first and second proteins are integral membrane proteins.

37. The fusion protein according to Claim 32 to 36 wherein the said first protein *is E. coli* cytochrome *bo*3.

38. The fusion protein according to Claim 37 wherein the said second protein is attached to subunit IV of *E. coli* cytochrome *bo*3.

39. A method for crystallization of a protein, comprising

(i) obtaining a fusion protein comprising (I) a first protein, which is a membrane protein or multisubunit protein and which upon crystallization yields crystals having available space in the lattice, so as to facilitate crystallization of a second protein; and (II) the said (second) protein to be crystallized; and

(ii) crystallizing the said fusion protein

wherein the resulting crystal is capable of diffracting x-rays.

40. A method for crystallization of a protein, comprising

(i) obtaining according to the process as defined in Claim 32, a fusion protein; and

(ii) crystallizing the said fusion protein

wherein the resulting crystal is capable of diffracting x-rays.

41. A method for crystallization of a protein, comprising

(i) obtaining a fusion protein as defined in any one of Claims 33 to 38; and

(ii) crystallizing the said fusion protein.

42. A method according to any one of Claims 39 to 41 wherein the first protein is an integral membrane protein.

43. A method for crystallization of a protein, comprising

(i) obtaining a first protein which is an integral membrane protein and which upon crystallization yields crystals having available space in the lattice so as to facilitate crystallization of a second protein; and

(ii) obtaining the second protein to be crystallized; and

(iii) crystallizing both the said proteins together

wherein the resulting crystal is capable of diffracting x-rays.

44. A method according to Claim 43 wherein the second protein is soaked into a crystal of the first protein.

45. A method according to any one of Claims 39 to 44 wherein the first protein is as defined in any one of Claims 1 to 24.

46. A method according to any one of Claims 39 to 45 further comprising a step wherein at least two detergents are screened in the crystal growth

conditions to identify which one optimizes the growth and/or diffraction of the resulting crystals.

47. A method according to any one of Claims 39 to 46 further comprising a step wherein the pH is optimized for crystal growth.

48. A method according to any one of Claims 39 to 47 wherein the crystal space group of the first protein when crystallized alone may be different to that obtained by crystallization with the second protein.

49. A method according to any one of Claims 39 to 48 wherein the second protein is an integral membrane protein.

50. A method according to any one of Claims 39 to 49 wherein the second protein has a lower molecular weight than the first protein.

51. A method of obtaining structural data on a protein of interest comprising the steps of

(i) obtaining the protein of interest;

(ii) crystallising said protein in the crystal lattice of another protein, which crystal lattice is able to accommodate the protein of interest; and

(iii) obtaining x-ray diffraction data from the crystal produced in step (ii).

52. A method according to Claim 51 wherein the crystallisation method is according to any one of Claims 39 to 49.

53. A method according to Claim 51 or 52 wherein the protein of interest is obtained by expressing a recombinant vector according to any one of Claims 26 to 29 or by culturing a cell according to Claim 30.

54. A method according to any one of Claims 51 to 53 wherein the protein of interest is an integral membrane protein.

55. A method according to any one of Claims 51 to 53 wherein the x-ray diffraction data is obtained to a resolution of at least 6Å.

56. Use of a recombinant vector according to any one of Claims 27 to 29 or a cell according to Claim 30 in a method according to any one of Claims 51 to 55.

57. A process for the production of a recombinant vector according to Claim 1 comprising

(i) obtaining a recombinant vector comprising (I) a nucleotide sequence encoding a first protein which is a membrane protein or multisubunit protein and which, when crystallized with a second protein, is capable of accommodating the second protein in the crystal lattice and (II) a promoter operably linked to the said nucleotide sequence; and

(ii) introducing, into the said vector, nucleotide sequences facilitating the insertion of further nucleotide sequences

wherein the resulting crystal would be capable of diffracting x-rays.

58. A process for the production of a recombinant vector according to Claim 3, comprising

(i) obtaining a recombinant vector comprising (I) a nucleotide sequence encoding a first protein which is a membrane protein or

multisubunit protein and which upon crystallization yields crystals having available space in the lattice, so as to allow for the ordered packing of a second protein into the said available space, and (II) a promoter operably linked to the said nucleotide sequence; and

(ii) introducing, into the said vector, nucleotide sequences facilitating the insertion of further nucleotide sequences

wherein the resulting crystal would be capable of diffracting x-rays.

59. The process according to Claim 57 or 58 wherein the said recombinant vector obtained in step (i) comprises the nucleotide sequence shown as SEQ ID NO: 1

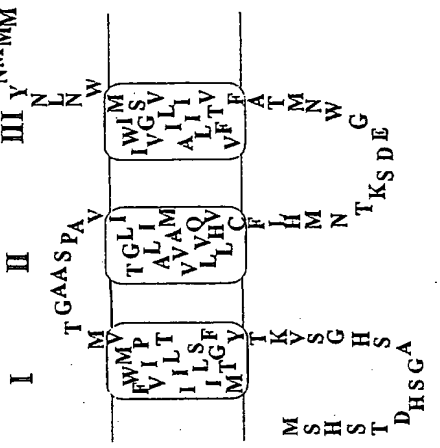


Figure 2

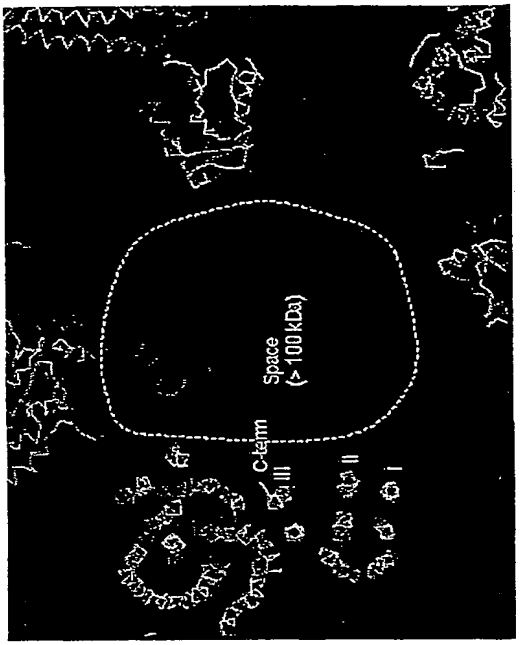
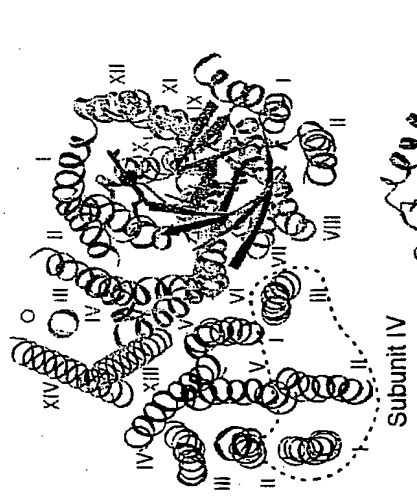


Figure 3



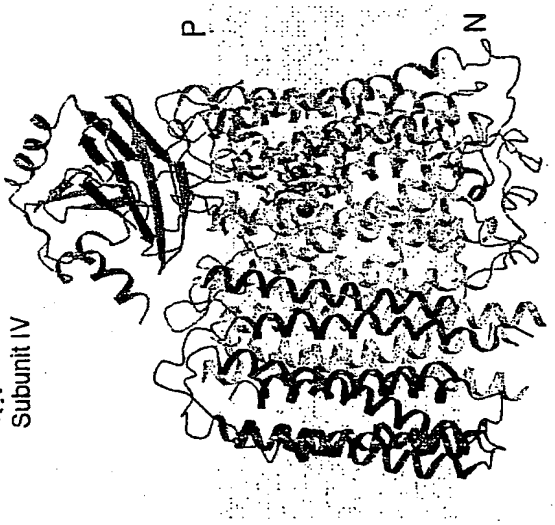


Figure 1

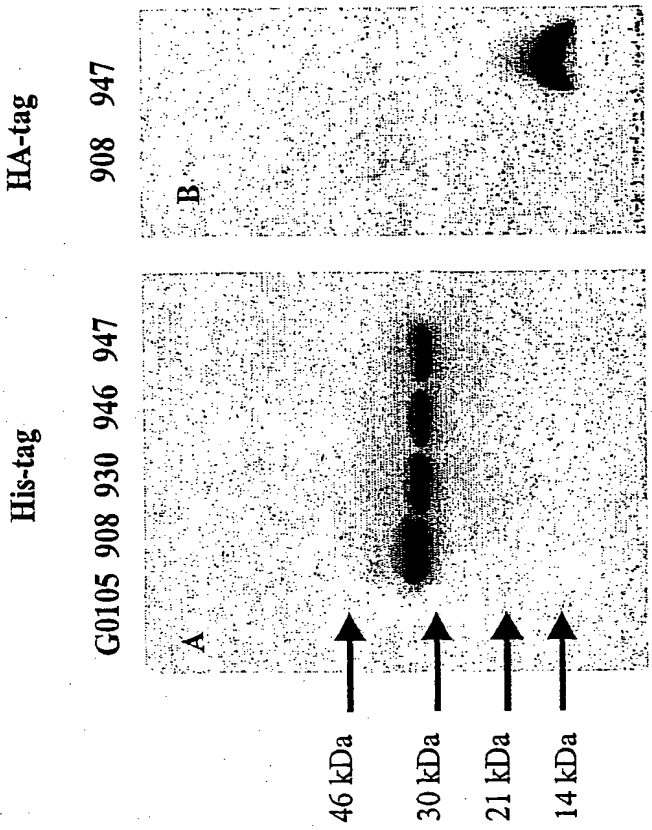


Figure 5

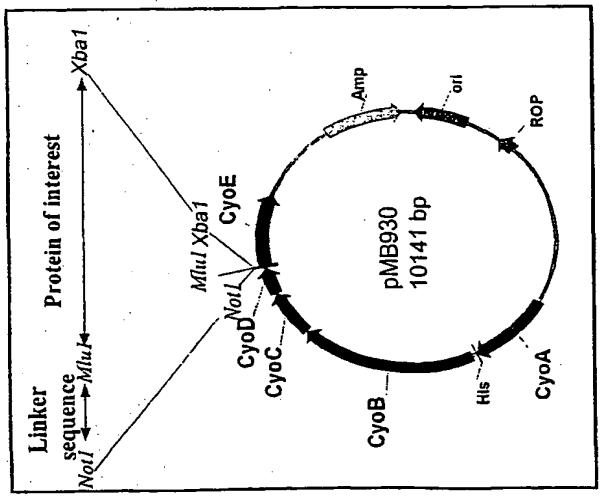


Figure 4

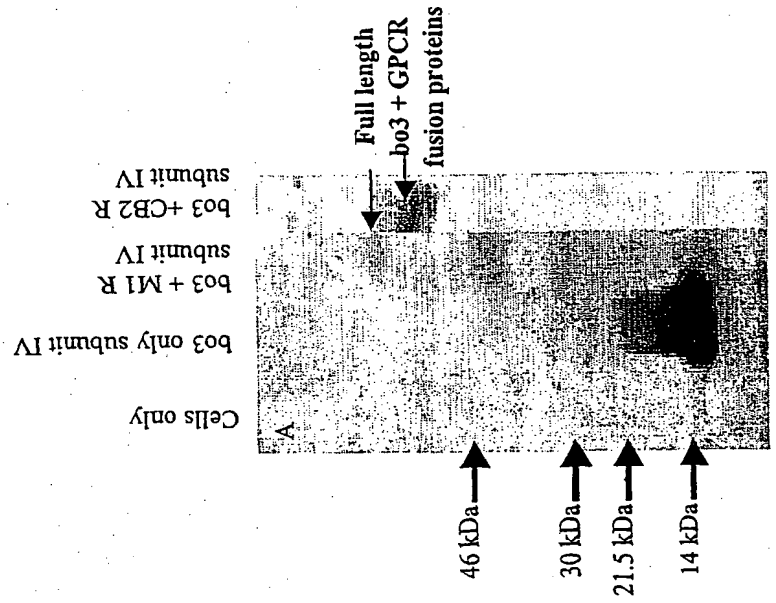


Figure 7

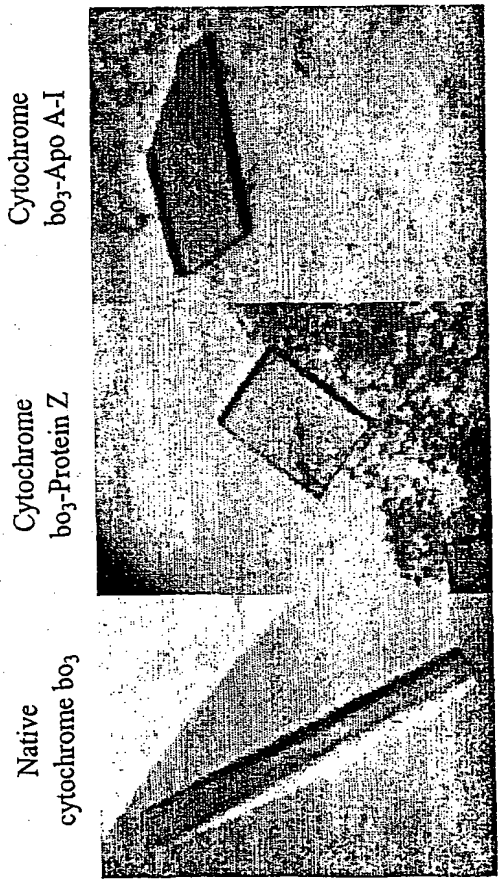


Figure 6

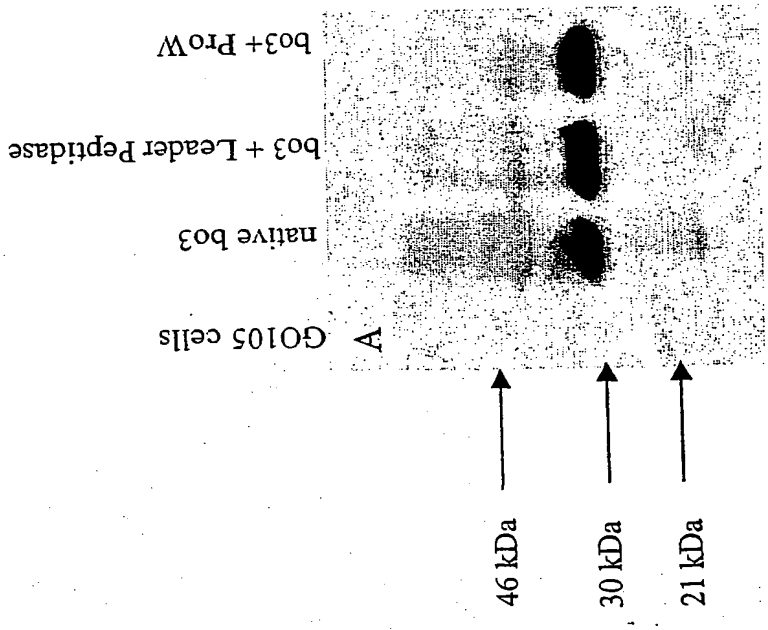

bo3+ ProW
bo3 + Leader Peptidase
native bo3
GO105 cells
A
46 kDa
30 kDa
21 kDa

Figure 8.

Figure 9

A
apo A-I fragment
908
1241
1242
1243
1244
Whole cells
1244
1241
1242
1243
1244
Membranes
49 kDa
38 kDa
18 kDa
B
GO105
908
1241
1242
1244
1244
Whole cells
1242
1241
1244
Membranes
46 kDa
30 kDa
21 kDa

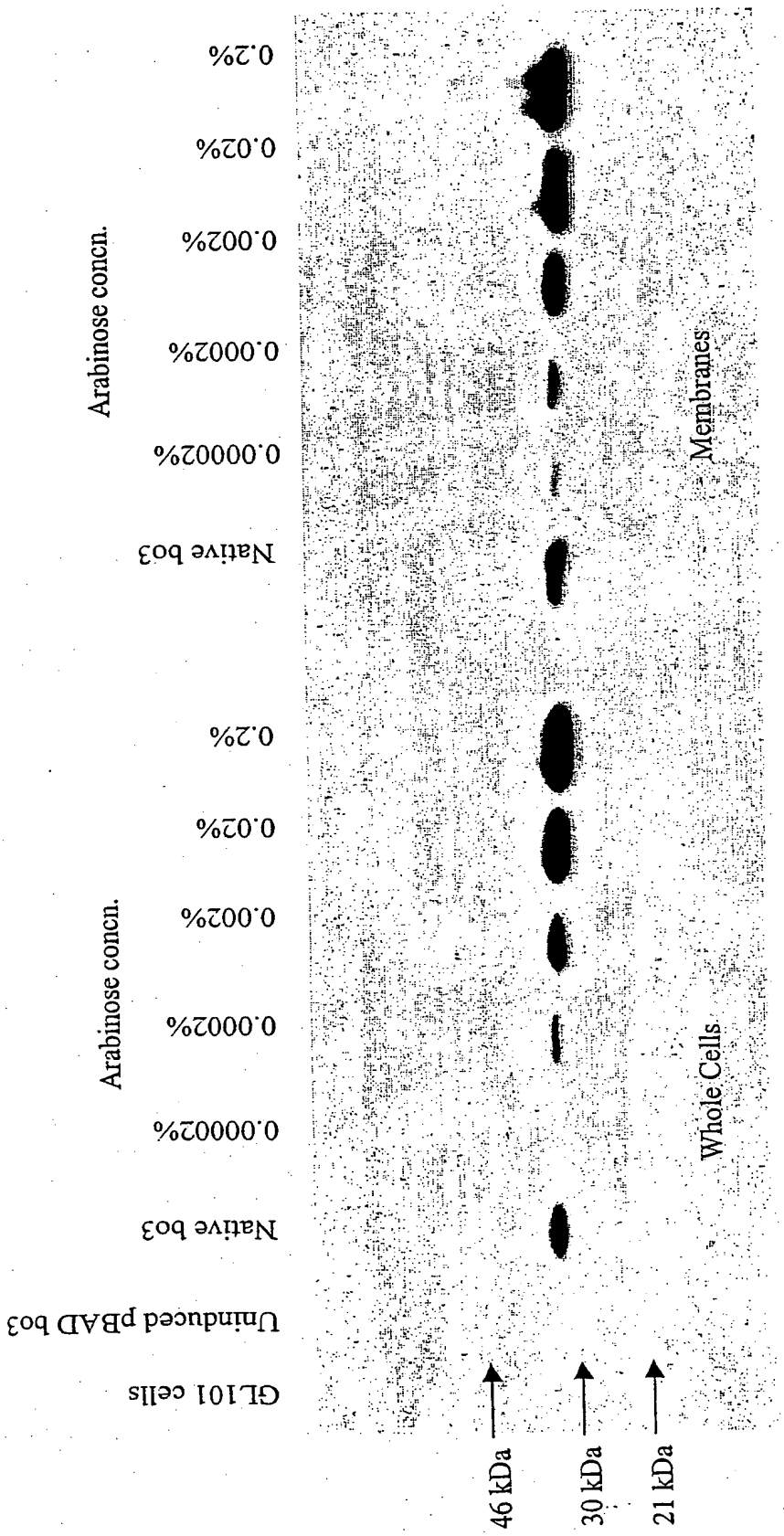

GL101 cells
Uninduced pBAD bo3
Native bo3
0.00002%
0.0002%
0.002%
0.02%
0.2%
Arabinose concn.
Native bo3
0.00002%
0.0002%
0.002%
0.02%
0.2%
Arabinose concn.
Whole Cells
Membranes
46 kDa
30 kDa
21 kDa

Figure 10

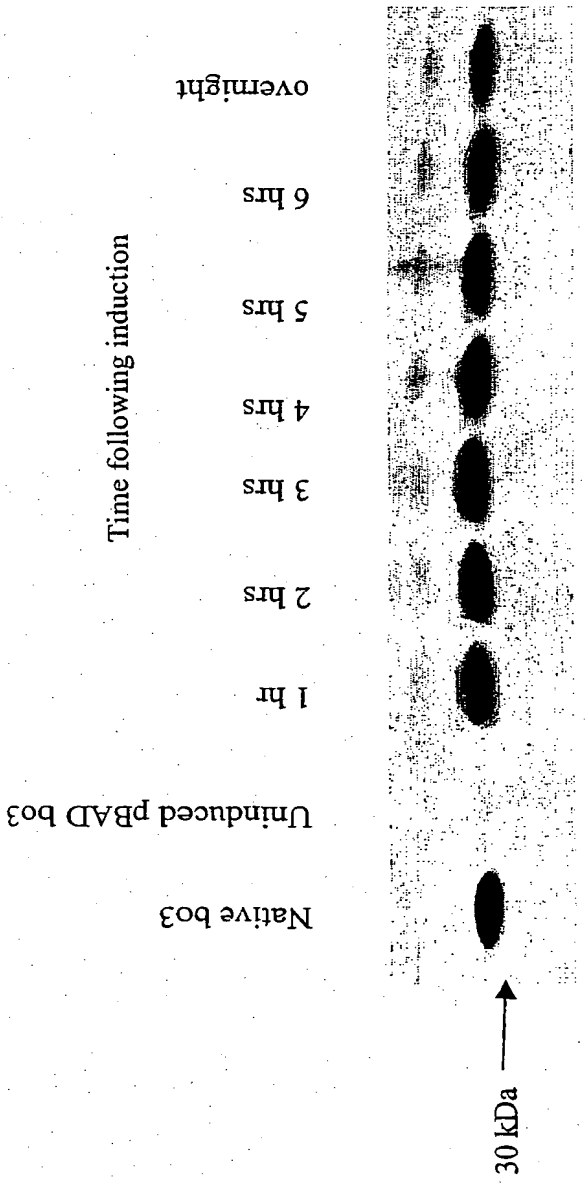
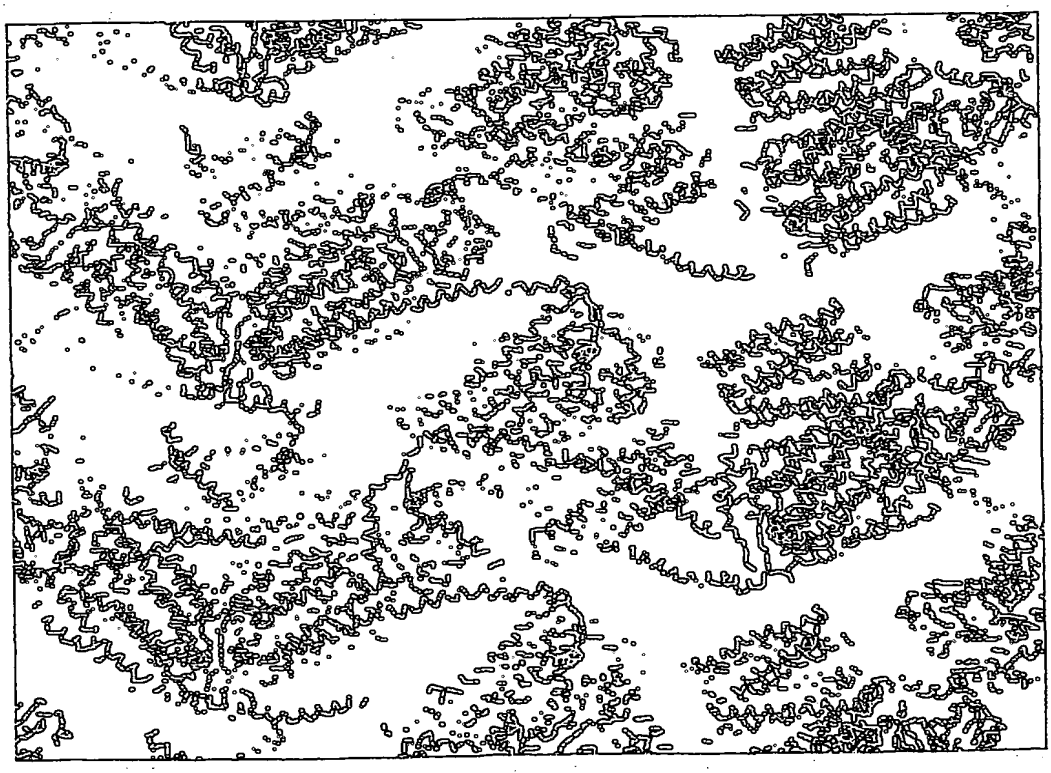

Native bo3
Uninduced pBAD bo3
1 hr
2 hrs
3 hrs
4 hrs
5 hrs
6 hrs
overnight
Time following induction
30 kDa

Figure 11

Figure 12



SEQUENCE LISTING

<110> Pharmacia & Upjohn AB

<120> Fusion vectors

<130> 00126

<140>
<141>

<160> 14

<170> PatentIn Ver. 2.1

<210> 1
<211> 10111
<212> DNA
<213> Artificial Sequence

<220>
<221> promoter
<222> (203)..(803)

<220>
<221> -35_signal
<222> (726)..(731)

<220>
<221> -10_signal
<222> (749)..(755)

<220>
<221> misc_feature
<222> (793)..(797)
<223> Shine-Dalgarno sequence

<220>
<221> misc_feature
<222> (804)..(1778)
<223> ORF coding for cyoA (bo3 subunit II) including His-tag

<220>
<221> misc_feature
<222> (1746)..(1772)
<223> His-tag

<220>
<221> misc_feature
<222> (1800)..(3791)
<223> ORF coding for cyoB (bo3 subunit I)

<220>
<221> misc_feature
<222> (3781)..(4395)
<223> ORF coding for cyoC (bo3 subunit III)

<220>
<221> misc_feature
<222> (4395)..(4724)

<223> ORF coding for cyoD (bo3 subunit IV)

<220>
<221> misc_feature
<222> (4736)..(5626)
<223> ORF coding for cyoE

<300>
<301> Rumbley, J.N.
<303> Biochim. Biophys. Acta
<304> 1340
<306> 131-142
<307> 1997
<313> 1 TO 10111

<300>
<301> Mingawa, J.
<303> J. Biol. Chem.
<304> 265
<305> 19
<306> 11198-11203
<307> 1990
<313> 203 TO 803

<400> 1
catcggtcga tcgacattct atctattctc cgtcgccgct gccgtaccag ggcttatttt 60
gctgctggtt tgccgccaga cgcttgaata tacacgagta aatgacaact ttatctcccg 120
taccgcatat ccggcaggtt atgcctttgc catgtggaca ctggcggcgg gcgtcagcct 180
gttggccgtg tggttactgc tattgacgat ggacgcgctg gatttgacgc acttctcttt 240
cctgcctgct ctgctggaag tcggggtttt agtcgccctt tctggcgtcg tgcttggtgg 300
tttgctggat tatctggcgc tacgaaaaac gcatctgacg taatctgtaa atattattta 360
gctgtagtaa tcactcgccg taattattac ggctaaataa acatcagtat cgcttattaa 420
tatttgttct tttgtgcggc ttagcgtttg gaaattattg gcgccattta taaattctat 480
ttttcttaca cgattcagct aatgagtctt tattttctca tcacccagtt gtcactctaa 540
tgataattat ttgttgaata attgttttat ttcacattgg ttataccaat tgcccgccca 600
gactccttca gcactcccct tttgttataa cgccctttg caacagcttc ttaaaatcaa 660
cctgatatgt tttgccaaca tatgtgacct ggcagccaaa tccaagtaac aggaatttaa 720
tcatgtttac agtaatttaa ccttcccgta aaatgcccac acactttaaa cgccaccaga 780
tcccgtggaa ttgaggtcgt taaatgagac tcaggaaata caataaaagt ttgggatggt 840
tgtcattatt tgcaggcact gtattgctca gtggctgtaa ttctgcgctg ttagatccca 900
aaggacagat tggtctggag caacgttcac tgatactgac ggcatttggc ctgatgttga 960
ttgtcgttat tcccgcaatc ttgatggctg ttggtttcgc ctggaagtac cgtgcgagca 1020
ataaagatgc taagtacagc ccgaactggt cacactccaa taaagtggaa gctgtggtct 1080
ggacggtacc tatcttaatc atcatcttcc ttgcagtact gacctggaaa accactcacg 1140
ctcttgagcc tagcaagccg ctggcacacg acgagaagcc cattaccatc gaagtggttt 1200
ccatggactg gaaatggttc ttcatctacc cggaacaggg cattgctacc gtgaatgaaa 1260
tcgctttccc ggcgaacact ccggtgtact tcaaagtgac ctccaactcc gtgatgaact 1320
ccttcttcat tccgcgtctg ggtagccaga tttatgccat ggccggtatg cagactcgcc 1380
tgcatctgat cgccaacgaa cccggcactt atgacggtat ctccgccagc tacagcggcc 1440
cgggcttctc aggcatgaag ttcaaagcta ttgcaacacc ggatcgcgcc gcattcgacc 1500
agtgggtcgc aaaagcgaag cagtcgccga acaccatgtc tgacatggct gcgttcgaaa 1560
aactggccgc gcctagcgaa tacaaccagg tggaatattt ctccaacgtg aaaccagact 1620
tgtttgccga tgtaattaac aagtttatgg ctcacggtaa gagcatggac atgacccagc 1680
cagaaggtga gcacagcgca cacgaaggta tggaaggcat ggacatgagc cacgcggaat 1740
ccgcccatca ccatcaccat caccatcacc atcgataaag gggttgagga agaataaaga 1800
tgttcggaaa attatcactt gatgcagtcc cgttccatga acctatcgtc atggttacga 1860
tcgctggcat tattttggga ggtctggcgc tcgttggcct gatcacttac ttcggtaagt 1920
ggacctacct gtggaaagag tggctgacct ccgtcgacca taaacgcctc ggtatcatgt 1980
atatcatcgt ggcgattgtg atgttgctgc gtggttttgc tgacgccatt atgatgcgta 2040
gccagcaggc tcttgcctcg gcgggcgaag cgggcttcct gccacctcac cactacgatc 2100
agatctttac cgcgcacggc gtgattatga tcttcttcgt agcgatgcct ttcgttatcg 2160

gtctgatgaa cctggtggtt ccgctgcaga tcggcgcgcg tgacgttgcg ttcccgttcc 2220
tcaacaactt aagcttctgg tttaccgttg ttggtgtgat tctggttaac gtttctctcg 2280
gcgtgggcga atttgcgcag accggctggc tggcctatcc accgctatcg ggaatagagt 2340
acagtccggg agtcggtgtc gattactgga tatggagtct ccagctatcc ggtataggta 2400
cgacgcttac cggtatcaac ttcttcgtta ccattctgaa gatgcgcgca ccgggcatga 2460
ccatgttcaa gatgccagta tttacctggg catcactgtg cgcgaacgta ctgattattg 2520
cttccttccc aattctgacg gttaccgtcg cgttgttgac cctggatcgc tatctgggca 2580
cccatttctt taccaacgat atgggtggca acatgatgat gtacatcaac ctgatttggg 2640
cctggggcca cccggaagtt tacatcctga tcctgcctgt tttcggtgtg ttctccgaaa 2700
ttgcggcaac cttctcgcgt aaacgtctgt ttggttatac ctcgctggta tgggcaaccg 2760
tctgtatcac cgtgctgtcg ttcatcgttt ggctgcacca cttctttacg atgggtgcgg 2820
gcgcgaacgt aaacgccttc tttggtatca ccacaatgat tatcgccatc ccgaccgggg 2880
tgaagatctt caactggctg ttcaccatgt atcagggccg catcgtgttc cattctgcga 2940
tgctgtggac catcggtttt atcgtcacct tctcggtggg cgggatgact ggcgtgctgc 3000
tggccgtacc gggcgcggac ttcgttctgc ataacagcct gttcctgatt gcgcacttcc 3060
ataacgtgat catcggcggc gtggtcttcg gctgcttcgc agggatgacc tactggtggc 3120
ctaaagcgtt cggtttcaaa ctgaacgaaa cctggggtaa acgcgcgttc tggttctgga 3180
tcatcggctt cttcgttgcc tttatgccac tgtatgcgct gggcttcatg ggcatgaccc 3240
gtcgtttgag ccagcagatt gacccgcagt tccacaccat gctgatgatt gcagccagcg 3300
gtgcagtact gattgcgctg ggtattctct gcctcgttat tcagatgtac gtttctattc 3360
gcgaccgcga ccagaaccgt gacctgactg gcgacccgtg gggtggccgt acgctggagt 3420
gggcaacctc ttccccgcct ccgttctata actttgccgt agtgccgcac gttcacgaac 3480
gtgatgcatt ctgggaaatg aaagagaaag gcgaagcgta taaaaagcct gaccactatg 3540
aagaaattca tatgccgaaa aacagcggtg caggtatcgt cattgcagct ttctccacca 3600
tcttcggttt cgccatgatc tggcatatct ggtggctggc gattgttggc ttcgcaggca 3660
tgatcatcac ctggatcgtg aagagcttcg acgaggacgt ggattactac gtgccggtgg 3720
cagaaatcga aaaactggaa aaccagcatt tcgatgagat tactaaggca gggctgaaaa 3780
atggcaactg atactttgac gcacgcgact gcccacgcgc acgaacacgg gcaccacgat 3840
gcaggcggaa ccaaaatctt cggattttgg atctacctga tgagcgactg cattctgttc 3900
tctatcttgt ttgctaccta tgccgttctg gtgaacggca ccgcaggcgg cccgacaggt 3960
aaggacattt tcgaactgcc gttcgttctg gttgaaactt tcttgctgtt gttcagctcc 4020
atcacctacg gcatggcggc tatcgccatg tacaaaaaca acaaaagcca ggttatctcc 4080
tggctggcgt tgacctggtt gtttggtgcc ggatttatcg ggatggaaat ctatgaattc 4140
catcacctga ttgttaacgg catgggtccg gatcgcagcg gcttcctgtc agcgttcttt 4200
gcgctggtcg gcacgcacgg tctgcacgtc acttctggtc ttatctggat ggcggtgctg 4260
atggtgcaaa tcgcccgtcg cggcctgacc agcactaacc gtacccgcat catgtgcctg 4320
agcctgttct ggcacttcct ggatgtggtt tggatctgtg tgttcactgt tgtttatctg 4380
atgggggcga tgtaatgagt cattctaccg atcacagcgg cgcgtcccat ggcagcgtaa 4440
aaacctacat gacaggcttt atcctgtcga tcattctgac ggtgattccg ttctggatgg 4500
tgatgacagg agctgcctct ccggccgtaa ttctgggaac aatcctggca atggcagtgg 4560
tacaggttct ggtgcatctg gtgtgcttcc tgcacatgaa taccaaatca gatgaaggct 4620
ggaacatgac ggcgtttgtc ttcaccgtgc taatcatcgc tatcctggtt gtaggctcca 4680
tctggattat gtggaacctc aactacaaca tgatgatgca ctaagagcgg cggttatgat 4740
gtttaagcaa tacctgcaag taacgaaacc aggcatcatc tttggcaacc tgatctcggt 4800
gattgggggа ttcctgctgg cctcaaaggg cagcattgat tatcccctgt ttatctacac 4860
gctggttggg gtgtcactgg ttgtggcgtc gggttgtgtg tttaacaact acatcgacag 4920
ggatatcgac agaaagatgg aaaggacgaa gaatcgggtg ctggtgaaag gcctgatctc 4980
tcctgctgtc tcgctggtgt acgccacgtt gctgggtatt gctggcttta tgctgctgtg 5040
gtttggcgcg aatccgctgg cctgctggct gggggtgatg ggctttgtgg tttatgtcgg 5100
cgtttatagc ctgtacatga aacgccactc tgtctacggc acgttgattg gttcgctctc 5160
cggcgctgcg ccgccggtga tcggctactg tgcggtaacc ggtgagttcg atagcggcgc 5220
agcgatcctg ctggctatct tcagcctgtg gcagatgcct cactcctatg ccatcgccat 5280
tttccgcttt aaggattacc aggcggcaaa cattccggta ttgccagtgg taaaaggcat 5340
ttcggtggcg aagaatcaca tcacgctgta tatcatcgcc tttgccgttg ccacgctgat 5400
gctctctctt ggcggttacg ctgggtataa atatctggtg gtcgccgcgg cggttagcgt 5460
ctggtggtta ggtatggctc tgcgcggtta taaagttgct gatgacagaa tctgggcgcg 5520
caagctgttc ggcttctcta tcatcgccat cactgccctc tcggtgatga tgtccgttga 5580
ttttatggta ccggactcgc atacgctgct ggctgctgtg tggtaacaaa acctctctat 5640
taaaaaggtg ctacggcacc ttttttctta gcattagaaa catatccctc tcgaaatatt 5700
tactaaaaaa tccgcatgtt taccccattc gtttgccgct ttacactagt cgcgaattta 5760
aaacagaggt ggtaatgaac gattataaaa tgacgccagg tgagaggcgc gcgacctggg 5820

gtttagggac cgtattctcg ttgcgcatgc aaggagatgg cgcccaacag tcccccggcc 5880
acggggcctg ccaccatacc cacgccgaaa caagcgctca tgagcccgaa gtggcgagcc 5940
cgatcttccc catcggtgat gtcggcgata taggcgccag caaccgcacc tgtggcgccg 6000
gtgatgccgg ccacgatgcg tccggcgtag aggatccaca ggacgggtgt ggtcgccatg 6060
atcgcgtagt cgatagtggc tccaagtagc gaagcgagca ggactgggcg gcggccaaag 6120
cggtcggaca gtgctccgag aacgggtgcg catagaaatt gcatcaacgc atatagcgct 6180
agcagcacgc catagtgact ggcgatgctg tcggaatgga cgatatcccg caagaggccc 6240
ggcagtaccg gcataaccaa gcctatgcct acagcatcca gggtgacggt gccgaggatg 6300
acgatgagcg cattgttaga tttcatacac ggtgcctgac tgcgttagca atttaactgt 6360
gataaactac cgcattaaag ctattcgatg ataagctgtc aaacatgagc attcttgaag 6420
acgaaagggc ctcgtgatac gcctattttt ataggttaat gtcatgataa taatggtttc 6480
ttagacgtca ggtggcactt ttcggggaaa tgtgcgcgga acccctattt gtttatttt 6540
ctaaatacat tcaaatatgt atccgctcat gagacaataa ccctgataaa tgcttcaata 6600
atattgaaaa aggaagagta tgagtattca acatttccgt gtcgccctta ttccctttt 6660
tgcggcattt tgccttcctg tttttgctca cccagaaacg ctggtgaaag taaaagatgc 6720
tgaagatcag ttgggtgcac gagtgggtta catcgaactg gatctcaaca gcggtaagat 6780
ccttgagagt tttcgccccg aagaacgttt tccaatgatg agcactttta aagttctgct 6840
atgtggcgcg gtattatccc gtgttgacgc cgggcaagag caactcggtc gccgcataca 6900
ctattctcag aatgacttgg ttgagtactc accagtcaca gaaaagcatc ttacggatgg 6960
catgacagta agagaattat gcagtgctgc cataaccatg agtgataaca ctgcggccaa 7020
cttacttctg acaacgatcg gaggaccgaa ggagctaacc gcttttttgc acaacatggg 7080
ggatcatgta actcgccttg atcgttggga accggagctg aatgaagcca taccaaacga 7140
cgagcgtgac accacgatgc ctgcagcaat ggcaacaacg ttgcgcaaac tattaactgg 7200
cgaactactt actctagctt cccggcaaca attaatagac tggatggagg cggataaagt 7260
tgcaggacca cttctgcgct cggcccttcc ggctggctgg tttattgctg ataaatctgg 7320
agccggtgag cgtgggtctc gcggtatcat tgcagcactg gggccagatg gtaagccctc 7380
ccgtatcgta gttatctaca cgacggggag tcaggcaact atggatgaac gaaatagaca 7440
gatcgctgag ataggtgcct cactgattaa gcattggtaa ctgtcagacc aagtttactc 7500
atatatactt tagattgatt taaaacttca tttttaattt aaaaggatct aggtgaagat 7560
ccttttgat aatctcatga ccaaaatccc ttaacgtgag ttttcgttcc actgagcgtc 7620
agaccccgta gaaaagatca aaggatcttc ttgagatcct ttttttctgc gcgtaatctg 7680
ctgcttgcaa acaaaaaaac caccgctacc agcggtggtt tgtttgccgg atcaagagct 7740
accaactctt tttccgaagg taactggctt cagcagagcg cagataccaa atactgtcct 7800
tctagtgtag ccgtagttag gccaccactt caagaactct gtagcaccgc ctacatacct 7860
cgctctgcta atcctgttac cagtggctgc tgccagtggc gataagtcgt gtcttaccgg 7920
gttggactca agacgatagt taccggataa ggcgcagcgg tcgggctgaa cgggggttc 7980
gtgcacacag cccagcttgg agcgaacgac ctacaccgaa ctgagatacc tacagcgtga 8040
gctatgagaa agcgccacgc ttcccgaagg gagaaaggcg gacaggtatc cggtaagcgg 8100
cagggtcgga acaggagagc gcacgaggga gcttccaggg ggaaacgcct ggtatcttta 8160
tagtcctgtc gggtttcgcc acctctgact tgagcgtcga tttttgtgat gctcgtcagg 8220
ggggcggagc ctatggaaaa acgccagcaa cgcggccttt ttacggttcc tggccttttg 8280
ctggcctttt gctcacatgt tctttcctgc gttatcccct gattctgtgg ataaccgtat 8340
taccgccttt gagtgagctg ataccgctcg ccgcagccga acgaccgagc gcagcgagtc 8400
agtgagcgag gaagcggaag agcgcctgat gcggtatttt ctccttacgc atctgtgcgg 8460
tatttcacac cgcatatggt gcactctcag tacaatctgc tctgatgccg catagttaag 8520
ccagtataca ctccgctatc gctacgtgac tgggtcatgg ctgcgccccg acacccgcca 8580
acacccgctg acgcgccctg acgggcttgt ctgctcccgg catccgctta cagacaagct 8640
gtgaccgtct ccgggagctg catgtgtcag aggttttcac cgtcatcacc gaaacgcgcg 8700
aggcagctgc ggtaaagctc atcagcgtgg tcgtgaagcg attcacagat gtctgcctgt 8760
tcatccgcgt ccagctcgtt gagtttctcc agaagcgtta atgtctggct tctgataaag 8820
cgggccatgt taagggcggt tttttcctgt ttggtcactg atgcctccgt gtaaggggga 8880
tttctgttca tgggggtaat gataccgatg aaacgagaga ggatgctcac gatacgggtt 8940
actgatgatg aacatgcccg gttactggaa cgttgtgagg gtaaacaact ggcggtatgg 9000
atgcggcggg accagagaaa aatcactcag ggtcaatgcc agcgcttcgt taatacagat 9060
gtaggtgttc cacagggtag ccagcagcat cctgcgatgc agatccggaa cataatggtg 9120
cagggcgctg acttccgcgt ttccagactt tacgaaacac ggaaaccgaa gaccattcat 9180
gttgttgctc aggtcgcaga cgttttgcag cagcagtcgc ttcacgttcg ctcgcgtatc 9240
ggtgattcat tctgctaacc agtaaggcaa ccccgccagc ctagccgggt cctcaacgac 9300
aggagcacga tcatgcgcac ccgtggccag gacccaacgc tgcccgagat gcgccgcgtg 9360
cggctgctgg agatggcgga cgcgatggat atgttctgcc aagggttggt ttgcgcattc 9420
acagttctcc gcaagaattg attggctcca attcttggag tggtgaatcc gttagcgagg 9480

tgccgccggc ttccattcag gtcgaggtgg cccggctcca tgcaccgcga cgcaacgcgg 9540
ggaggcagac aaggtatagg gcggcgccta caatccatgc caacccgttc catgtgctcg 9600
ccgaggcggc ataaatcgcc gtgacgatca gcggtccagt gatcgaagtt aggctggtaa 9660
gagccgcgag cgatccttga agctgtccct gatggtcgtc atctacctgc ctggacagca 9720
tggcctgcaa cgcgggcatc ccgatgccgc cggaagcgag aagaatcata atggggaagg 9780
ccatccagcc tcgcgtcgcg aacgccagca agacgtagcc cagcgcgtcg gccgccatgc 9840
cggcgataat ggcctgcttc tcgccgaaac gtttggtggc gggaccagtg acgaaggctt 9900
gagcgagggc gtgcaagatt ccgaataccg caagcgacag gccgatcatc gtcgcgctcc 9960
agcgaaagcg gtcctcgccg aaaatgaccc agagcgctgc cggcacctgt cctacgagtt 10020
gcatgataaa gaagacagtc ataagtgcgg cgacgatagt catgccccgc gcccaccgga 10080
aggagctgac tgggttgaag gctctcaagg g 10111

<210> 2
<211> 30
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: PCR primer

<400> 2
ttcacgaacg tgatgcattc tgggaaatga 30

<210> 3
<211> 30
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: PCR primer

<400> 3
ctccttgcat gcgcaagcag aatacggtcc 30

<210> 4
<211> 52
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: PCR primer

<400> 4
gcacagcggc cgcgagctca cgcgtatgtc tagataagag cggcggttat ga 52

<210> 5
<211> 49
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: PCR primer

<400> 5
atctagacat acgcgtgagc tcgcggccgc tgtgcatcat catgttgta 49

<210> 6
<211> 9
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Strep-tag linker

<400> 6
Ala Trp Arg His Pro Gln Phe Gly Gly
1 5

<210> 7
<211> 34
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Strep-tag linker, forward strand

<400> 7
ggccgcgcct ggcgtcatcc tcaattcggt ggca 34

<210> 8
<211> 34
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Strep-tag linker, reverse strand

<400> 8
cgcgtgccaa attgaggatg acgccgccag gcgc 34

<210> 9
<211> 18
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Strep-HA-tag linker

<400> 9
Ala Trp Arg His Pro Gln Phe Gly Gly Tyr Pro Tyr Asp Val Pro Asp
1 5 10 15

Tyr Ala

<210> 10
<211> 9
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: HA-tag linker

<400> 10
Tyr Pro Tyr Asp Val Pro Asp Tyr Ala
1 5

<210> 11
<211> 61
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Strep-HA-tag linker, forward strand

<400> 11
ggccgcgcct ggcgtcatcc tcaattcggt ggctacccat acgacgtccc agactacgct 60
a 61

<210> 12
<211> 61
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Strep-HA-tag linker, reverse strand

<400> 12
cgcgtagcgt agtctgggac gtcgtatggg tagccaccga attgaggatg acgccaggcg 60
c 61

<210> 13
<211> 5819
<212> DNA
<213> E. coli

<220>
<221> misc_feature
<222> (801)..(1748)
<223> ORF coding for cyoA (bo3 subunit II)

<220>
<221> misc_feature
<222> (1770)..(3761)
<223> ORF coding for cyoB (bo3 subunit I)

<220>
<221> misc_feature
<222> (3751)..(4365)
<223> ORF coding for cyoC (bo3 subunit III)

<220>
<221> misc_feature
<222> (4365)..(4694)
<223> ORF coding for cyoD (bo3 subunit IV)

<220>
<221> misc_feature
<222> (4706)..(5596)
<223> ORF coding for cyoE

<300>
<301> Chepuri, V.
<303> J. Biol. Chem.
<304> 265
<305> 19
<306> 11185-11192
<307> 1990
<308> GenBank/J05492
<313> 1 TO 5819

<400> 13
actactgtcg acattctatc tattctccgt cgccgctgcc gtaccagggc ttattttgct 60
gctggtttgc cgccagacgc ttgaatatac acgagtaaat gacaacttta tctcccgtac 120
cgcatatccg gcaggttatg cctttgccat gtggacactg gcggcgggcg tcagcctgtt 180
ggccgtgtgg ttactgctat tgacgatgga cgcgctggat ttgacgcact tctctttcct 240
gcctgctctg ctggaagtcg gggttttagt cgcccttttct ggcgtcgtgc ttggtggttt 300
gctggattat ctggcgctac gaaaaacgca tctgacgtaa tctgtaaata ttatttagct 360
gtagtaatca ctcgccgtaa ttattacggc taaataaaca tcagtatcgc ttattaatat 420
ttgttctttt gtgcggctta gcgtttggaa attattggcg ccatttataa attctatttt 480
tcttacacga ttcagctaat gagtctttat tttctcatca cccagttgtc actctaatga 540
taattatttg ttgaataatt gttttatttc acattggtta taccaattgc ccgcccagac 600
tccttcagca ctccccttt gttataacgc ccttttgcaa cagcttctta aaatcaacct 660
gatatgtttt gccaacatat gtgacctggc agccaaatcc aagtaacagg aatttaatca 720
tgtttacagt aatttaacct tcccgtaaaa tgcccacaca ctttaaacgc caccagatcc 780
cgtggaattg aggtcgttaa atgagactca ggaaatacaa taaaagtttg ggatggttgt 840
cattatttgc aggcactgta ttgctcagtg gctgtaattc tgcgctgtta gatcccaaag 900
gacagattgg tctggagcaa cgttcactga tactgacggc atttggcctg atgttgattg 960
tcgttattcc cgcaatcttg atggctgttg gtttcgcctg gaagtaccgt gcgagcaata 1020
aagatgctaa gtacagcccg aactggtcac actccaataa agtggaagct gtggtctgga 1080
cggtacctat cttaatcatc atcttccttg cagtactgac ctggaaaacc actcacgctc 1140
ttgagcctag caagccgctg gcacacgacg agaagcccat taccatcgaa gtggtttcca 1200
tggactggaa atggttcttc atctacccgg aacagggcat tgctaccgtg aatgaaatcg 1260
ctttcccggc gaacactccg gtgtacttca aagtgacctc caactccgtg atgaactcct 1320
tcttcattcc gcgtctgggt agccagattt atgccatggc cggtatgcag actcgcctgc 1380
atctgatcgc caacgaaccc ggcacttatg acggtatctc cgccagctac agcggcccgg 1440
gcttctcagg catgaagttc aaagctattg caacaccgga tcgcgccgca ttcgaccagt 1500
gggtcgcaaa agcgaagcag tcgccgaaca ccatgtctga catggctgcg ttcgaaaaac 1560
tggccgcgcc tagcgaatac aaccaggtgg aatatttctc caacgtgaaa ccagacttgt 1620
ttgccgatgt aattaacaag tttatggctc acggtaagag catggacatg acccagccag 1680
aaggtgagca cagcgcacac gaaggtatgg aaggcatgga catgagccac gcggaatccg 1740
cccattaaag gggttgagga agaataaaga tgttcggaaa attatcactt gatgcagtcc 1800
cgttccatga acctatcgtc atggttacga tcgctggcat tattttggga ggtctggcgc 1860
tcgttggcct gatcacttac ttcggtaagt ggacctacct gtggaaagag tggctgacct 1920
ccgtcgacca taaacgcctc ggtatcatgt atatcatcgt ggcgattgtg atgttgctgc 1980
gtggttttgc tgacgccatt atgatgcgta gccagcaggc tcttgcctcg gcgggcgaag 2040
cgggcttcct gccacctcac cactacgatc agatctttac cgcgcacggc gtgattatga 2100
tcttcttcgt agcgatgcct ttcgttatcg gtctgatgaa cctggtggtt ccgctgcaga 2160
tcggcgcgcg tgacgttgcg ttcccgttcc tcaacaactt aagcttctgg tttaccgttg 2220
ttggtgtgat tctggttaac gtttctctcg gcgtgggcga atttgcgcag accggctggc 2280
tggcctatcc accgctatcg ggaatagagt acagtccggg agtcggtgtc gattactgga 2340
tatggagtct ccagctatcc ggtataggta cgacgcttac cggtatcaac ttcttcgtta 2400
ccattctgaa gatgcgcgca ccgggcatga ccatgttcaa gatgccagta tttacctggg 2460
catcactgtg cgcgaacgta ctgattattg cttccttccc aattctgacg gttaccgtcg 2520
cgttgttgac cctggatcgc tatctgggca cccatttctt taccaacgat atgggtggca 2580
acatgatgat gtacatcaac ctgatttggg cctggggcca cccggaagtt tacatcctga 2640

tcctgcctgt tttcggtgtg ttctccgaaa ttgcggcaac cttctcgcgt aaacgtctgt 2700
ttggttatac ctcgctggta tgggcaaccg tctgtatcac cgtgctgtcg ttcatcgttt 2760
ggctgcacca cttctttacg atgggtgcgg gcgcgaacgt aaacgccttc tttggtatca 2820
ccacaatgat tatcgccatc ccgaccgggg tgaagatctt caactggctg ttcaccatgt 2880
atcagggccg catcgtgttc cattctgcga tgctgtggac catcggtttt atcgtcacct 2940
tctcggtggg cgggatgact ggcgtgctgc tggccgtacc gggcgcggac ttcgttctgc 3000
ataacagcct gttcctgatt gcgcacttcc ataacgtgat catcggcggc gtggtcttcg 3060
gctgcttcgc agggatgacc tactggtggc ctaaagcgtt cggtttcaaa ctgaacgaaa 3120
cctggggtaa acgcgcgttc tggttctgga tcatcggctt cttcgttgcc tttatgccac 3180
tgtatgcgct gggcttcatg ggcatgaccc gtcgtttgag ccagcagatt gacccgcagt 3240
tccacaccat gctgatgatt gcagccagcg gtgcagtact gattgcgctg ggtattctct 3300
gcctcgttat tcagatgtac gtttctattc gcgaccgcga ccagaaccgt gacctgactg 3360
gcgacccgtg gggtggccgt acgctggagt gggcaacctc ttccccgcct ccgttctata 3420
actttgccgt agtgccgcac gttcacgaac gtgatgcatt ctgggaaatg aaagagaaag 3480
gcgaagcgta taaaaagcct gaccactatg aagaaattca tatgccgaaa aacagcggtg 3540
caggtatcgt cattgcagct ttctccacca tcttcggttt cgccatgatc tggcatatct 3600
ggtggctggc gattgttggc ttcgcaggca tgatcatcac ctggatcgtg aaaagcttcg 3660
acgaggacgt ggattactac gtgccggtgg cagaaatcga aaaactggaa aaccagcatt 3720
tcgatgagat tactaaggca gggctgaaaa atggcaactg atactttgac gcacgcgact 3780
gcccacgcgc acgaacacgg gcaccacgat gcaggcggaa ccaaaatctt cggattttgg 3840
atctacctga tgagcgactg cattctgttc tctatcttgt ttgctaccta tgccgttctg 3900
gtgaacggca ccgcaggcgg cccgacaggt aaggacattt tcgaactgcc gttcgttctg 3960
gttgaaactt tcttgctgtt gttcagctcc atcacctacg gcatggcggc tatcgccatg 4020
tacaaaaaca acaaaagcca ggttatctcc tggctggcgt tgacctggtt gtttggtgcc 4080
ggatttatcg ggatggaaat ctatgaattc catcacctga ttgttaacgg catgggtccg 4140
gatcgcagcg gcttcctgtc agcgttcttt gcgctggtcg gcacgcacgg tctgcacgtc 4200
acttctggtc ttatctggat ggcggtgctg atggtgcaaa tcgcccgtcg cggcctgacc 4260
agcactaacc gtacccgcat catgtgcctg agcctgttct ggcacttcct ggatgtggtt 4320
tggatctgtg tgttcactgt tgtttatctg atgggggcga tgtaatgagt cattctaccg 4380
atcacagcgg cgcgtcccat ggcagcgtaa aaacctacat gacaggcttt atcctgtcga 4440
tcattctgac ggtgattccg ttctggatgg tgatgacagg agctgcctct ccggccgtaa 4500
ttctgggaac aatcctggca atggcagtgg tacaggttct ggtgcatctg gtgtgcttcc 4560
tgcacatgaa taccaaatca gatgaaggct ggaacatgac ggcgtttgtc ttcaccgtgc 4620
taatcatcgc tatcctggtt gtaggctcca tctggattat gtggaacctc aactacaaca 4680
tgatgatgca ctaagagcgg cggttatgat gtttaagcaa tacctgcaag taacgaaacc 4740
aggcatcatc tttggcaacc tgatctcggt gattggggga ttcctgctgg cctcaaaggg 4800
cagcattgat tatcccctgt ttatctacac gctggttggg gtgtcactgg ttgtggcgtc 4860
gggttgtgtg tttaacaact acatcgacag ggatatcgac agaaagatgg aaaggacgaa 4920
gaatcgggtg ctggtgaaag gcctgatctc tcctgctgtc tcgctggtgt acgccacgtt 4980
gctgggtatt gctggcttta tgctgctgtg gtttggcgcg aatccgctgg cctgctggct 5040
gggggtgatg ggctttgtgg tttatgtcgg cgtttatagc ctgtacatga aacgccactc 5100
tgtctacggc acgttgattg gttcgctctc cggcgctgcg ccgccggtga tcggctactg 5160
tgcggtaacc ggtgagttcg atagcggcgc agcgatcctg ctggctatct tcagcctgtg 5220
gcagatgcct cactcctatg ccatcgccat tttccgcttt aaggattacc aggcggcaaa 5280
cattccggta ttgccagtgg taaaaggcat ttcggtggcg aagaatcaca tcacgctgta 5340
tatcatcgcc tttgccgttg ccacgctgat gctctctctt ggcggttacg ctgggtataa 5400
atatctggtg gtcgccgcgg cggttagcgt ctggtggtta ggtatggctc tgcgcggtta 5460
taaagttgct gatgacagaa tctgggcgcg caagctgttc ggcttctcta tcatcgccat 5520
cactgccctc tcggtgatga tgtccgttga ttttatggta ccggactcgc atacgctgct 5580
ggctgctgtg tggtaacaaa acctctctat taaaaaggtg ctacggcacc ttttttctta 5640
gcattagaaa catatccctc tcgaaatatt tactaaaaaa tccgcatgtt taccccattc 5700
gtttgccgct ttacactagt cgcgaattta aaacagaggt ggtaatgaac gattataaaa 5760
tgacgccagg tgagaggcgc gcgacctggg gtttagggac cgtattctcg ttgcgcatg 5819

<210> 14
<211> 58
<212> PRT
<213> Staphylococcus aureus

<400> 14

```
                                                    10
Val Asp Asn Lys Phe Asn Lys Glu Gln Gln Asn Ala Phe Tyr Glu Ile
 1               5                  10                  15

Leu His Leu Pro Asn Leu Asn Glu Glu Gln Arg Asn Ala Phe Ile Gln
            20                  25                  30

Ser Leu Lys Asp Asp Pro Ser Gln Ser Ala Asn Leu Leu Ala Glu Ala
        35                  40                  45

Lys Lys Leu Asn Asp Ala Gln Ala Pro Lys
    50                  55
```

INTERNATIONAL SEARCH REPORT

Int[redacted]ation No
P[redacted]02043

A. CLASSIFICATION OF SUBJECT MATTER
IPC 7 C12N15/62 C12N15/70 C07K1/30

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 7 C12N C07K

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)
EPO-Internal, WPI Data, PAJ, BIOSIS, CHEM ABS Data, EMBASE

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	PRIVÉ, G.G. ET AL.: "Fusion Proteins as Tools for Crystallization: the Lactose Permease from Escherichia coli" ACTA CRYSTALLOGRAPHICA, vol. D50, no. 4, 1994, pages 375-379, XP001015503 page 375, right-hand column, paragraph 3 -page 376, left-hand column, paragraph 2 --- -/--	1-59

[X] Further documents are listed in the continuation of box C.

[ ] Patent family members are listed in annex.

° Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
1 August 2001	17/08/2001

Name and mailing address of the ISA	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax: (+31-70) 340-3016	Mata Vicente, T.

Form PCT/ISA/210 (second sheet) (July 1992)

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication,where appropriate, of the relevant passages	Relevant to claim No.
X	TURNER, G. J. ET AL.: "Expression, Purification, and Structural Characterization of the Bacteriorhodopsin Aspartyl Transcarbamylase Fusion Protein" PROTEIN EXPRESSION AND PURIFICATION, vol. 17, 1999, pages 324-338, XP002173694 page 325, left-hand column, paragraph 3 -right-hand column, paragraph 1 page 325, right-hand column, paragraph 3 -page 326, left-hand column, paragraph 1 ---	1-11, 16-18, 26-36, 39-58
P,X	BYRNE, B. ET AL.: "Fusion protein approach to improve the crystal quality of cytochrome bo3 ubiquinol oxidase from Escherichia coli" BIOCHIMICA ET BIOPHYSICA ACTA, vol. 1459, no. 2-3, 15 August 2000 (2000-08-15), pages 449-455, XP001010670 abstract -----	1-59